1-1-2013

N-Terminal Protein Modification by Aminoacyl Transferase: Enzymology and Applications.

Anne Wagner
University of Pennsylvania, annewagner05@gmail.com

N-Terminal Protein Modification by Aminoacyl Transferase: Enzymology and Applications.

Abstract
Protein modification is an important tool for understanding how a protein moves, folds, and operates in its cellular environment. We aim for a minimalist approach to protein modification in order to learn more about the protein while not perturbing its native state. Our method for ligation of small molecules to proteins has high N-terminal specificity and creates natural amide bond linkages under mild conditions. We utilized E. coli aminoacyl L/F transferase (AaT) in our ligation reaction to recognize a small synthetic substrate analog, which mimics aminoacylated tRNA, and can be chemically synthesized in only three steps. We have studied the mechanism of AaT in order better utilize unnatural substrates and expand the scope of aminoacyl donors and N-terminal acceptors. Using AaT, we can enzymatically attach a wide range of synthetic moieties to the N-termini of peptides and proteins, including natural and unnatural amino acids with reactive azido groups, fluorescent labels, protected disulfides, and photoinducible crosslinkers. This method introduces new opportunities for protein semi-synthesis, which we have applied in two model systems, α-synuclein and calmodulin, and contributes as an effective method to label proteins for biophysical research.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemistry

First Advisor
E. James Petersson

Keywords
amide ligation, semi-synthesis, transferase

Subject Categories
Biochemistry | Chemistry

N-TERMINAL PROTEIN MODIFICATION BY AMINOACYL TRANSFERASE: ENZYMOLOGY AND APPLICATIONS.

Anne M. Wagner

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

_____

E. James Petersson

Assistant Professor of Chemistry

Graduate Group Chairperson

_____

Gary A. Molander, Hirschmann-Makineni Professor of Chemistry

Dissertation Committee

Jeffery G. Saven, Associate Professor of Chemistry

Donna Huryn, Adjunct Professor of Chemistry

Ronen Marmorstein, Hilary Koprowski, M.D. Professor & Program Leader, Gene Expression and Regulation Program

N-TERMINAL PROTEIN MODIFICATION BY AMINOACYL TRANSFERASE:

ENZYMOLOGY AND APPLICATIONS.

DEDICATION

I want to dedicate this dissertation to all of the teachers, professors, and mentors that I have been fortunate to work with throughout my studies.

ACKNOWLEDGEMENTS

First, I would like to thank my research advisor, E. James Petersson. He has been a guiding mentor and very supportive throughout my PhD education. He has provided valuable analysis and advice for my current research and future career goals. His willingness to help has been greatly appreciated especially throughout the process of dissertation writing. I would also like to thank my committee members, Dr. Jeffery Saven, Dr. Donna Huryn, Dr. Ronen Marmorstein, and Dr. David Christianson who have helped keep my research focused and productive with their constructive suggestions and guidance.

There are many people in the chemistry department that have helped me have access to high quality scientific equipment and provided a workspace for me to thrive as a scientist. Thank you to all our directors of research facilities: Rakesh Koli for mass spectroscopy, George Furst and Jun Gu for NMR, and Christopher Lanci for the BCRC communal equipment. The administrative staff, James Tarver, Mandy Swope, and Candice Adams have helped me throughout the years to make sure I have all of my paperwork in order and a room reserved whenever I need it. Also, the stockroom employees have been a continual help throughout the years, thank you Cuong Nguyen and Andrei Korchynsky. I would also like to thank all of my research collaborators, Dr. Jeffery Saven, Lin Gu, Dr. Christopher MacDermaid, Dr. Anna Kashina, and Dr. Junling Wang.

I am also thankful to be a part of a great and hardworking group in the Petersson Lab. Thank you to all current and past members that I have had the pleasure of working with: Benson Chen, Stella Chen, Colin Fadzen, Mark Fegley, Haviva Garrett, Jacob Goldberg, Chrissie Grindley, Yun Huang, E. Keith Keenan, Nicholas Marotta, Eileen Moison, Anand Muthusamy, Lee C. Speight, Dr. Moumita Samanta, Dr. Tomohiro Tanaka, Chris Walters, Jerri Wang, John Warner, Rebecca Wissner, and Solongo Ziraldo. Thank you. I'd especially like to thank my EJP girls Rebecca and Solongo; you both have been so supportive and made spending long days in the lab a lot of fun with plenty of laughs. Thank you also to Jacob Goldberg for sitting at the desk next to me for five years and introducing me to a bunch of new music. We had a lot of fun playing "The End of the Line" by the Traveling Wilburys on repeat; it took the lab about five repeats for them to notice. I also want to let all of the EJP Labbers that have worked directly on my project know that I have appreciated all of their help, research, and useful advice: Mark Fegley, Haviva Garrett, Nicholas Marotta, Eileen Moison, Dr. Tomohiro Tanaka, Chris Walters, and John Warner. Good luck to Chris who will be taking over my project when I leave.

I also want to thank classmates and friends I have made over the years at Penn chemistry including Jasmina Cheung-Lau, Julie Griepenburg, Glen Liszczak, Genette McGrew, Chris MacDermaid, Robert Phillips, Brittany Riggle, Kate Smith, Paul Strauss, and Kate Thorn. Jasmina, you were my first friend at UPenn and helped me get through a tough first year. Julie, thank you for being so supportive and giving me the perspective I needed throughout the PhD process. We certainly had a lot of fun during our lunches

together. Thank you for being a great friend. Also, my bio group cohort really helped me with classes; thank you Jacob, Kate and Glen for making the studying and hard work fun.

I have been lucky to find some amazing kindred spirits in Philadelphia. Thank you to Natie Meindhart and Vonni Moorman for becoming my "Philly Family." You both have been such great friends and unwavering in your support and kindness. As one of Vonni's friends had said, "This (pointing at us three) should have never have happened." I'm thankful to have met you both and had so many fun times throughout the years.

For being a constant source of support, I want to thank my parents, David and Nancy Malhowski, and my twin sister, Amy Malhowski. Also, thank you to my inlaws, Eric and Taffy Wagner, and my sister-in-law Sophie Wagner. You all have been patient, confident in me, and given a willing ear when I've needed it. I'd also like to thank my dog, Fezzik, for being constantly loving and happy.

Lastly, I would like to thank my best friend and wonderful husband, David Wagner. You have always believed in me and had confidence in my abilities even when I have had difficulty seeing it in myself. You are a constant source of support, confidence, encouragement, and love. You helped me get through all the tough times in graduate school. I couldn't have done it without you to support me.

ABSTRACT

N-TERMINAL PROTEIN MODIFICATION BY AMINOACYL TRANSFERASE: ENZYMOLOGY AND APPLICATIONS.

Anne M. Wagner

E. James Petersson

Protein modification is an important tool for understanding how a protein moves, folds, and operates in its cellular environment. We aim for a minimalist approach to protein modification in order to learn more about the protein while not perturbing its native state. Our method for ligation of small molecules to proteins has high N-terminal specificity and creates natural amide bond linkages under mild conditions. We utilized *E. coli* aminoacyl L/F transferase (AaT) in our ligation reaction to recognize a small synthetic substrate analog, which mimics aminoacylated tRNA, and can be chemically synthesized in only three steps. We have studied the mechanism of AaT in order better utilize unnatural substrates and expand the scope of aminoacyl donors and N-terminal acceptors. Using AaT, we can enzymatically attach a wide range of synthetic moieties to the N-termini of peptides and proteins, including natural and unnatural amino acids with reactive azido groups, fluorescent labels, protected disulfides, and photoinducible crosslinkers. This method introduces new opportunities for protein semi-synthesis, which

we have applied in two model systems, α-synuclein and calmodulin, and contributes as an effective method to label proteins for biophysical research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION TO LEUCYL/PHENYLALANYL AMINOACYL TRANSFERASE AND CURRENT N-TERMINAL MODIFICATION METHODS

Section 1.1 – Introduction to Protein Engineering

If DNA is considered the "blueprint" of life, then an enzyme would aptly be described as the machine of life. Enzymes access a countless number of reactions to maintain cell function, and their biological activities have evolved to incredible levels of complexity over 4 billion years.[1] Natural products can be accessed through enzymatic reactivity that is mostly unavailable to synthetic chemists on the timescale and yield of the enzyme.[2] Learning more about the structure and function of enzyme molecular machines gives us insight into how enzymes can access their reactive abilities. One way to learn more about enzymes is through the process of protein engineering (Figure 1-1). By making primary sequence changes to an enzyme, you can change its original function.[3] However, haphazard mutational changes without thought of the molecular effects of the change are unlikely to be beneficial to enzymatic catalysis; evolution is successful because it is possible to select for very rare enhancements of function.[4] There are several methods that can be used to enhance the effectiveness of primary sequence mutations: rational design, computer-aided design, directed evolution, and protein semi-synthesis.[5] All of the aforementioned techniques can attain the engineered product, a new and improved protein, with a variety of possible alterations such as improved enzyme activity, altered specificity for a substrate or product, or enzyme regulation; they can also be used as a tool to understand inherent protein characteristics that remain mysterious, such as protein folding mechanisms.[6] The desired ultimate change in the protein is directly linked to the first step in changing the protein sequence. You can take protein engineering one step further by putting in an unnatural amino acid into the protein

in order to give it entirely new functionality that is unavailable with the 20 naturally occurring amino acids. In this introduction, I will review current methods of protein modification and delve into the heart of my thesis research with a background and history of leucyl/phenylalanyl aminoacyl transferase, which the Petersson group has utilized as an N-terminally modifying enzyme that can contribute to the field of protein engineering.

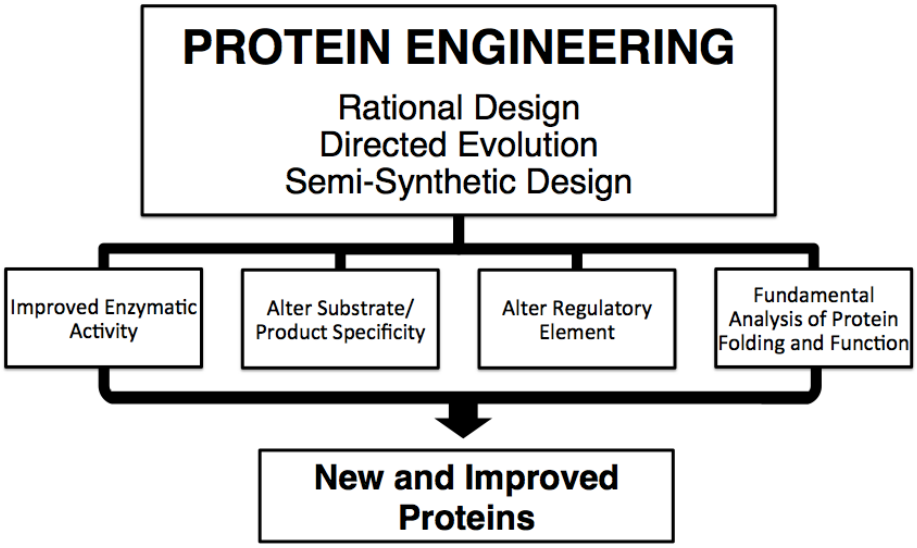


Figure 1-1. Protein Engineering Pathway
This flow chart describes modes of protein engineering and their respective outcomes adapted from Foo *et al*.[7]

Section 1.2 – Current Methods of Protein Modification

1.2.1 – Introduction to Amide Bond-Forming Ligation Techniques

Modifications of proteins in a site-selective manner can introduce a chemical consequence from which one can infer information about protein structure and function. Standard manipulations of protein structure include molecular biology techniques such amino acid mutagenesis.[8] However, these modifications are limited to natural amino acids, ribosomal synthesis, and uncontrolled posttranslational modifications.[9] One

current method for adding a synthetic group to a protein is solid-phase peptide synthesis (SPPS), which generates peptides *in vitro* (Figure 1-2). This method allows for sequence-specific addition of natural and unnatural amino acids. However, there are some limitations to this method; SPPS requires lengthy synthetic steps, the use of protecting groups, and undesirable peptide byproducts due to incomplete reactions.[10] The accumulation of peptide byproducts limits the number of amino acids that can be synthesized directly on solid phase with an upper limit at ~ 50 amino acids.[9] Despite its limitations, SPPS was and continues to be an advantageous method, and proteins, such as, RNAseA, have been fully synthesized by this method.[11] Through the use of SPPS, unnatural amino acids and backbone alterations, such as a thioamide, can be incorporated easily into peptides.[12]

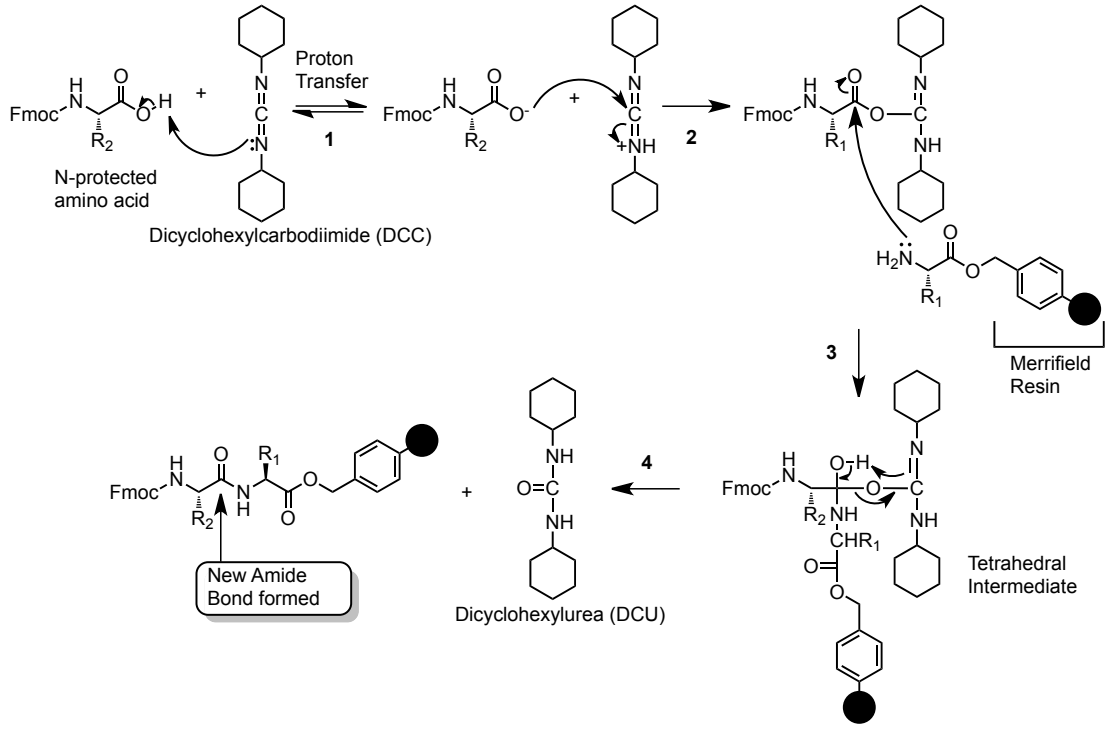


Figure 1-2. Mechanism of SPPS Amino Acid Activation and Amide Bond Formation
(1) DCC deprotonates (often with added base) the unprotected carboxylic acid of an N-terminal Fmoc-protected amino acid. (2) Negatively charged oxygen from the carboxylic acid attacks the central carbon of DCC, forming an activated amino acid. (3) The activated amino acid can now be attacked by an N-terminal amino acid that is resin bound. (4) The tetrahedral intermediate collapses to form a peptide bond. Steps 1-4 are repeated until desired length of peptide is reached. Peptide synthesis is completed with deprotection of all the amino acid protecting groups and cleavage from the solid-phase resin. R_1 and R_2 are protected amino acid sidechains.[13-14]

1.2.2 – Native Chemical Ligation

One technique that has expanded upon SPPS and can circumvent the limitation of peptide length is native chemical ligation (NCL). (Figure 1-3) NCL utilizes the reaction of a cysteine and thioester to generate an amide bond in aqueous conditions with near-physiological pH. Kent *et al.* used NCL between two large SPPS fragments, 51 and 48 amino acids in length, to create the create the monomer of the HIV-1 protease, one peptide containing a C-terminal thioester and the other fragment with an N-terminal

cysteine.[15] This ligation strategy was a breakthrough into studying larger proteins and without the need of amino acid protecting groups.[16] Two key aspects of NCL are its regioselectivity and chemoselectivity. In the presence of an exogenous thiol catalyst, NCL is regioselective because the thiol(ate)-thioester exchange is reversible.[9, 17] The NCL reaction demonstrates chemoselectivity because the reaction will only occur at a peptide or protein terminus and not at internal residues.[18] Thioesters have unique chemical properties in comparison to oxoesters, which contribute to their efficacy in NCL. Thioesters are more stable than oxoesters to hydroxide-catalyzed hydrolysis, but conversely, are also more reactive than oxoesters to thiolysis and aminolysis.[9] Thus, the thioester yields itself well to working with proteins and peptides in the NCL reaction because under physiological pH, the thioester will react readily with cysteine but not hydrolyze under these conditions.[9] In addition to its mild conditions, racemization does not occur at either the N or C-terminal residues of the connecting peptides. The aforementioned attributes of the NCL reaction make it a very powerful tool for protein engineering.

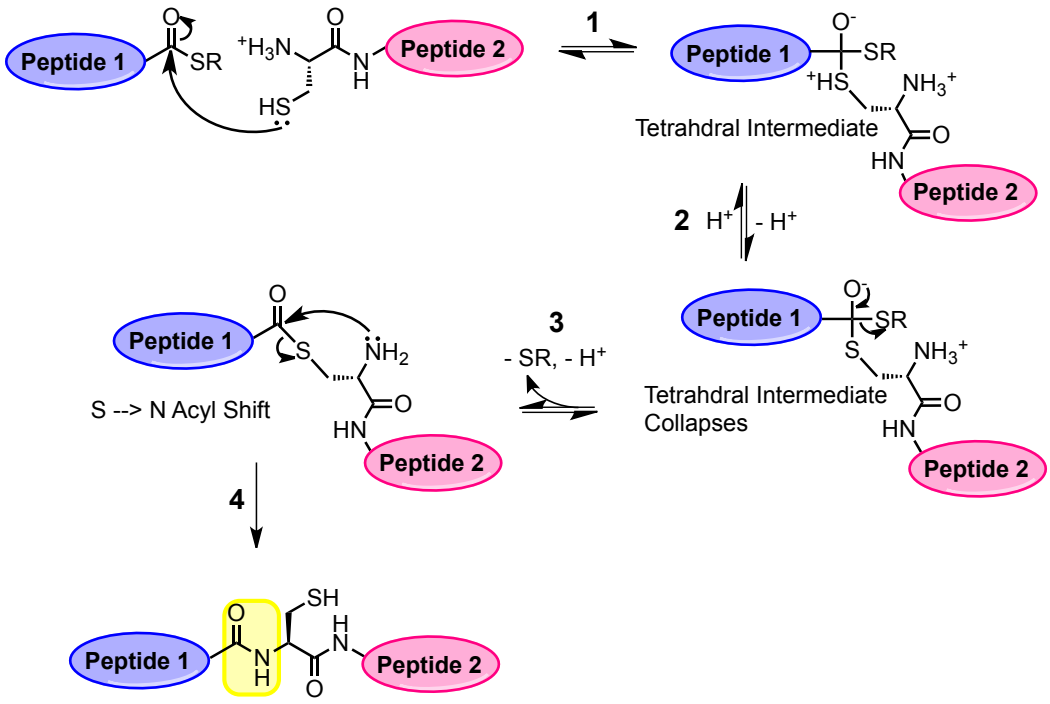


Figure 1-3. NCL Mechanism
(1) In the first step, the cysteine thiol attacks the c-terminal thioester carbonyl group. (2 and 3) A transthioesterification reaction occurs, forming a tetrahedral intermediate that collapses with expulsion of the SR leaving group. (4) An S to N acyl shift occurs, which is an irreversible reaction, forming the final product, peptide 1 and 2 ligated together *via* a peptide bond, highlighted in yellow.

1.2.3 – Expressed Protein Ligation

The NCL reaction has opened up the ability to synthetically create proteins that are not limited by the 50 amino acid restriction of length and can produce protein in near quantitative yield.[9] Additionally, Muir and coworkers have improved upon the NCL reaction through the use of inteins to make recombinant proteins that are functionalized with a C-terminal thioester with a method termed expressed protein ligation (EPL). EPL is a semi-synthetic technique that utilizes a recombinant protein with a fused intein, which can be thiolysed to form a C-terminal thioester *in situ* and then ligated with the N-terminal cysteine of a synthetic peptide or recombinant protein.[19] An intein is a protein

sequence within a precursor protein that is self-excised during the maturation process of a protein; the protein segments that are rejoined after intein excision are called exteins (Figure 1-4).[20-22] To make inteins a more useful tool for semi-synthesis, Muir and coworkers have made several modifications to native intein splicing. As seen in Figure 1-5, the EPL reaction and intein excision mechanism begin the same way, with an S-to-N acyl shift. However, for EPL, the intein has been mutated, preventing the continuation of the splicing mechanism, and the exogenous thiol, such as sodium 2-mercaptoethanesulfonate (MESNA), attacks the carbonyl of the transient thioester formed from the acyl shift in excision of the intein. An additional mutation at $N_{454}A$ prevents the intein from excising itself from its C-terminus. The C-terminus is fused to a chitin binding domain (CBD) or a His tag, which is used to purify out the thiolysed intein from the reaction. Then, with the addition of a C-terminal cysteine-bearing peptide or recombinant protein, the NCL reaction can proceed. EPL's exploitation of intein thiolysis allowed for the formation of C-terminal thioesters *in situ* on a recombinant protein. This procedure was a significant advancement in protein semi-synthesis because the protein fragments did not need to be synthesized using SPPS, and thus much larger fragments could be generated.

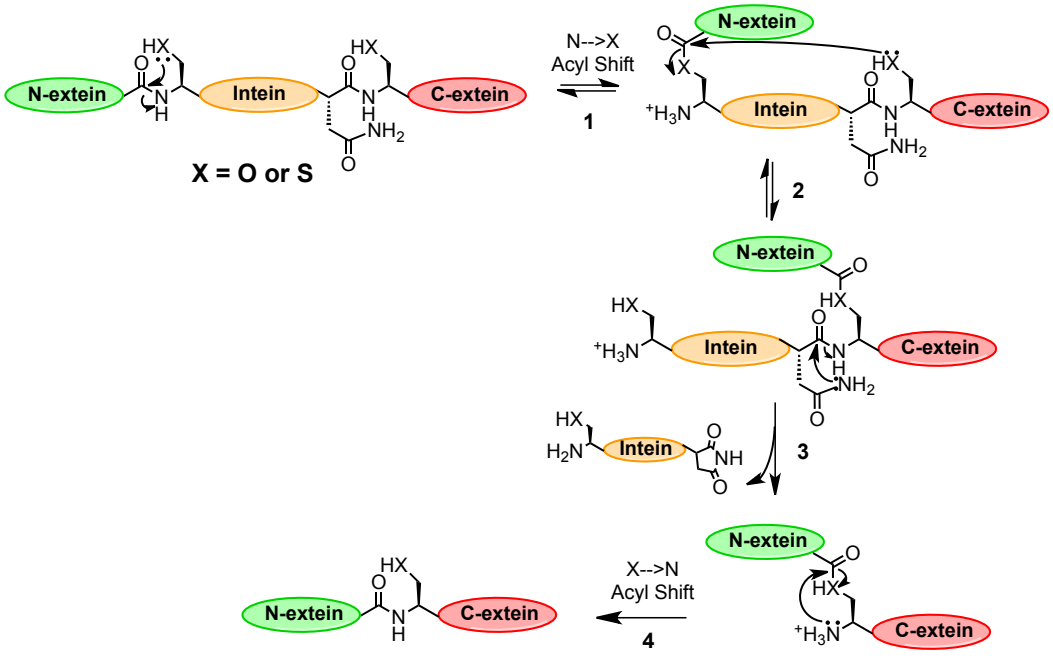


Figure 1-4. Mechanism of Intein Excision
(1) The protein splicing mechanism begins with an N-to-X acyl shift. X can be either an oxygen or sulfur atom. Now the N-extein has moved to the Cys/Ser residue. (2) Transthioesterification reaction occurs between a conserved Cys/Ser/Thr residue at the C-terminus cleavage site of the intein. (3) A conserved asparagine excises the intein from the two exteins. (4) The mechanism ends with a final X-to-N acyl shift, forming an amide bond.[23]

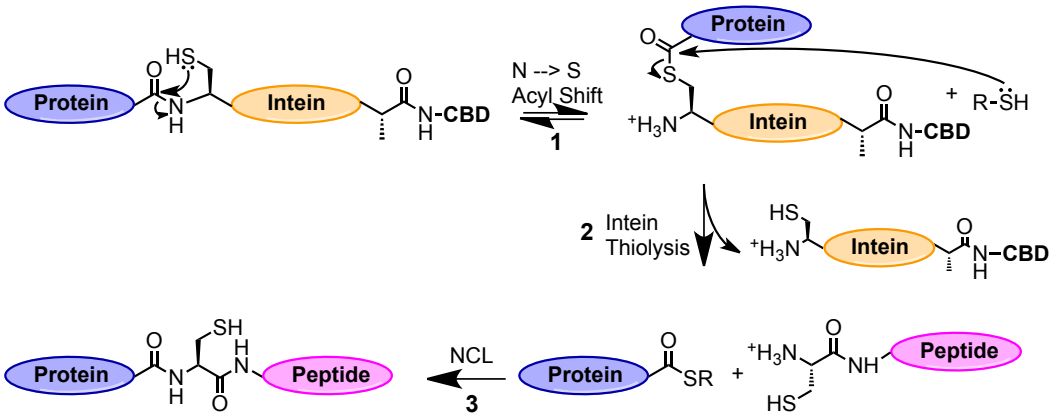


Figure 1-5. EPL Mechanism
(1) N-to-S acyl shift. (2) Transthioesterification resulting in intein thiolysis. (3) A thioester is formed from excision of the intein and can now perform NCL.[23] The R group of RSH can be either alkyl, phenyl, benzyl, or $CH_2CH_2SO_3$ (MESNA).[19] CBD = chitin binding domain.

1.2.4 – Traceless Staudinger Ligation

While NCL and EPL have helped to advance protein semi-synthesis, both reactions require a cysteine residue at the ligation site. This amino acid requirement can be a hindrance because the natural occurrence of cysteine in proteins is very low at less than 2 % with tryptophan being the least abundant.[9, 24] The low abundance of natural cysteine limits the opportunities for ligation sites that maintain original primary sequence. Additionally, cysteine mutagenesis can be detrimental to protein structure and function due to its nucleophilic and oxidative reactivities.[24] One strategy that circumvents the need for a cysteine residue is the traceless Staudinger ligation (TSL); this ligation reaction does not have any inherent residue restrictions at the ligation site (Figure 1-6). The TSL reaction scheme begins with transthioesterification between a phosphinothiol and a thioester-bearing peptide.[25] Now the first peptide is ready for reaction with the second peptide, which bears an N-terminal α-azido acid. The phosphorous and azide react to form an iminophosphorane, which can undergo subsequent rearrangements.[24] The negatively-charged nitrogen of the iminophosphorane will attack the carbonyl carbon of the thioester and form an amidophosphonium salt, which upon hydrolysis, yields the ligated peptide product and phosphine oxide byproduct.[16, 24-25] The original ligation was done in the presence of DMF, and since then a variety of phosphinothiol coupling reagents have been synthesized that have a range of ligation efficiencies and water solubilities.[26] The TSL reaction using water soluble phosphinothiols has been shown to be compatible with EPL so that C-terminal fragments can be prepared from expressed proteins.[26] Additionally, the TSL reaction has been modified by Bertozzi *et al*. through metabolic labeling using biotinylated water soluble phosphinothiol and metabolically-

incorporated azido-bearing sugars.[27-28] The TSL reaction has some limitations; in protein semi-synthesis the incorporation of the azido group still must be synthetically incorporated, and in addition, the phosphine can oxidize in the presence of air. Steric bulk at the ligating residues can also slow the reaction. The TSL reaction, even with these limitations, offers an alternate route for protein semi-synthesis at amino acids other than cysteine.[27]

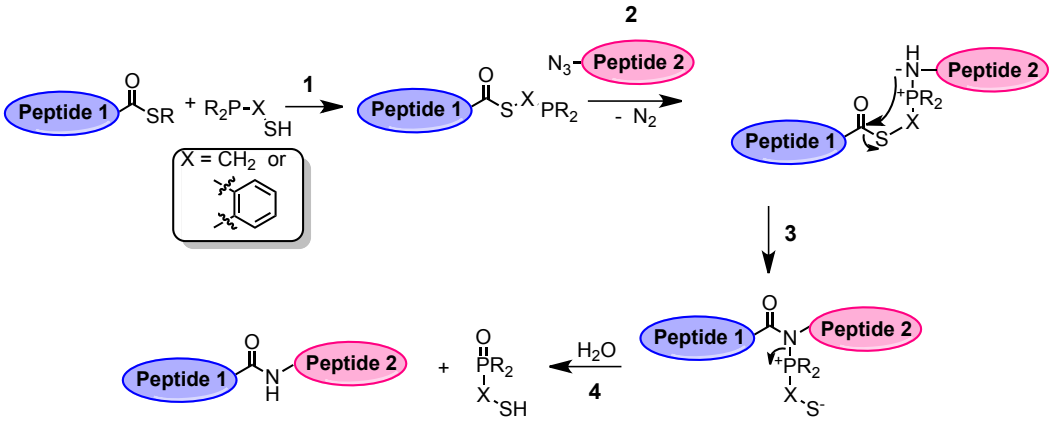


Figure 1-6. Traceless Staudinger Ligation Reaction
(1) Transthioesterification. (2) Formation of iminophosphorane. (3) Formation of amidophosphonium salt. (4) Hydrolysis of amidophosphonium salt forms amide bond.[15-16, 25]

1.2.5 – KAHA Ligation: α-Ketoacid–hydroxylamine amide-forming ligation reaction

Another ligating reaction, recently developed by Bode *et al*., is a chemoselective amide ligation occurring *via* decarboxylative condensation between an *N*-alkylhydroxylamine and α-ketoacid named the KAHA ligation. KAHA ligations are successful under two conditions, Type I and Type II, which differ by hydroxylamine substitution and solvent conditions (Figure 1-7). Type II *O*-substituted hydroxylamines have faster reaction kinetics in aqueous conditions, and Type I unsubstituted hydroxylamines are much slower and are superior in organic solvents.[29] Additionally,

Type II ligations require a lower hydroxylamine concentration for a successful ligation. The Bode group has investigated the KAHA ligation over several years and have proposed several mechanisms.[29-30] Most recently in 2011, Pusterla *et al*. elucidated the mechanism for Type I KAHA ligations by using ^{18}O labeling to determine which intermediates were involved in producing the final amide product (Figure 1-8). In this mechanism, formation of a nitrone is required, as opposed to being a non-productive byproduct as they initially proposed in their 2006 mechanism. They also speculated that the Type II ligation could follow direct oxidative decarboxylation as described in their initial 2006 mechanism or through an oxaziridine intermediate.[29-30] These ligations can be performed in conditions that maintain protein folding and stability; however, currently, successful ligations have only been performed on the peptide scale.

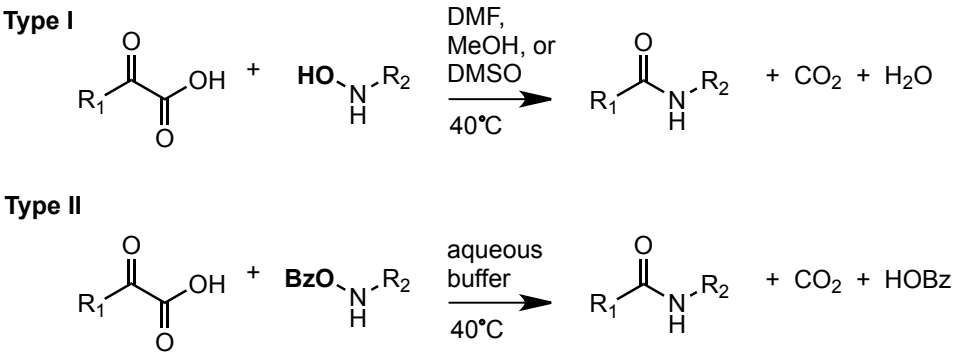


Figure 1-7. KAHA Ligation Types Depending on O-Substitution
Type I ligations use unsubstituted hydroxylamines and Type II ligations use substituted hydroxylamines.[29]

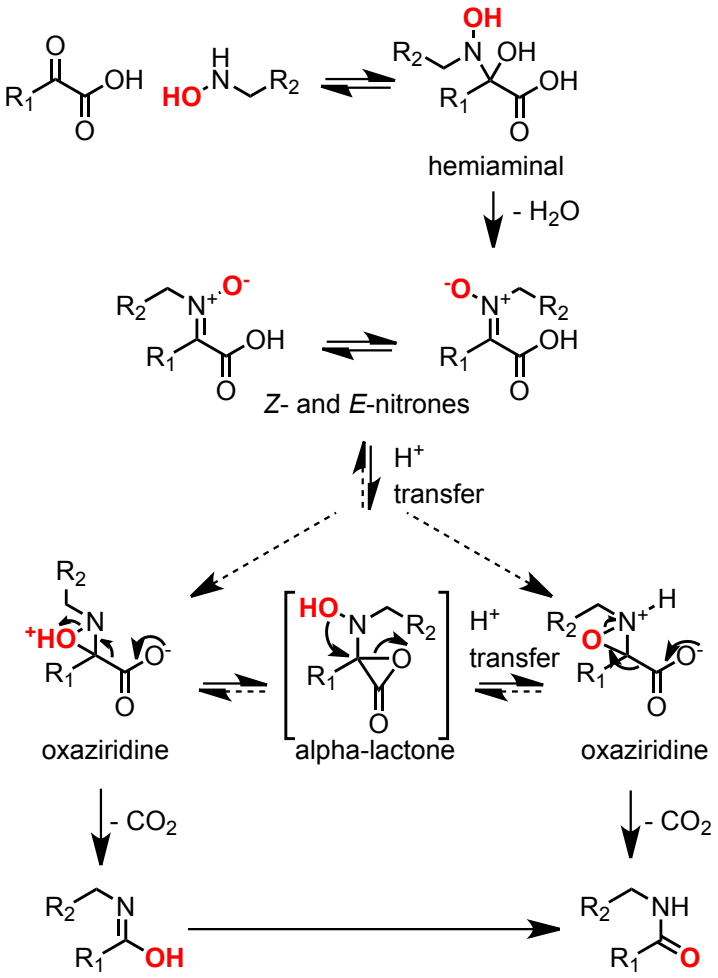


Figure 1-8. Bode Mechanism Proposal for Type I KAHA Ligation
It is proposed that the ligation proceeds through an alpha-lactone and oxaziridine intermediate.

1.2.6 – Reverse Proteolysis - Subtiligase

Reverse proteolysis is another option for protein ligation. Proteases have two possible reaction outcomes of the acyl intermediate, amide bond formation and proteolytic cleavage, which is the primary reaction.[31] One protease in particular, subtilisin, has been mutated to favor amide bond formation. In its wild type form, subtilisin has a k_{cat} that is approximately 10^9 greater than spontaneous amide cleavage. However, upon mutation of two residues, $S_{221}C$ and $P_{225}A$, a high efficiency protein ligase was created, called subtiligase (Figure 1-9). The $S_{221}C$ residue mutation is one of the catalytic triad amino acids and the cysteine mutation creates steric crowding during the formation of the oxyanion transition state required for cleavage, therefore, pushing the reaction towards ligation. Additionally, it is suspected that the $S_{221}C$ mutation slows hydrolysis and aids in bond formation. The $P_{225}A$ mutation aids in opening the active site to fit the $S_{221}C$ mutation. This double mutant has a greater than 10^7 reduction in amidase activity and a 500 fold increase in peptide bond formation preference over hydrolysis. Subtiligase is useful because it can catalyze peptide bond formation under protein-friendly conditions, but this reaction requires high concentrations of protein to reverse the reaction equilibrium.[32]

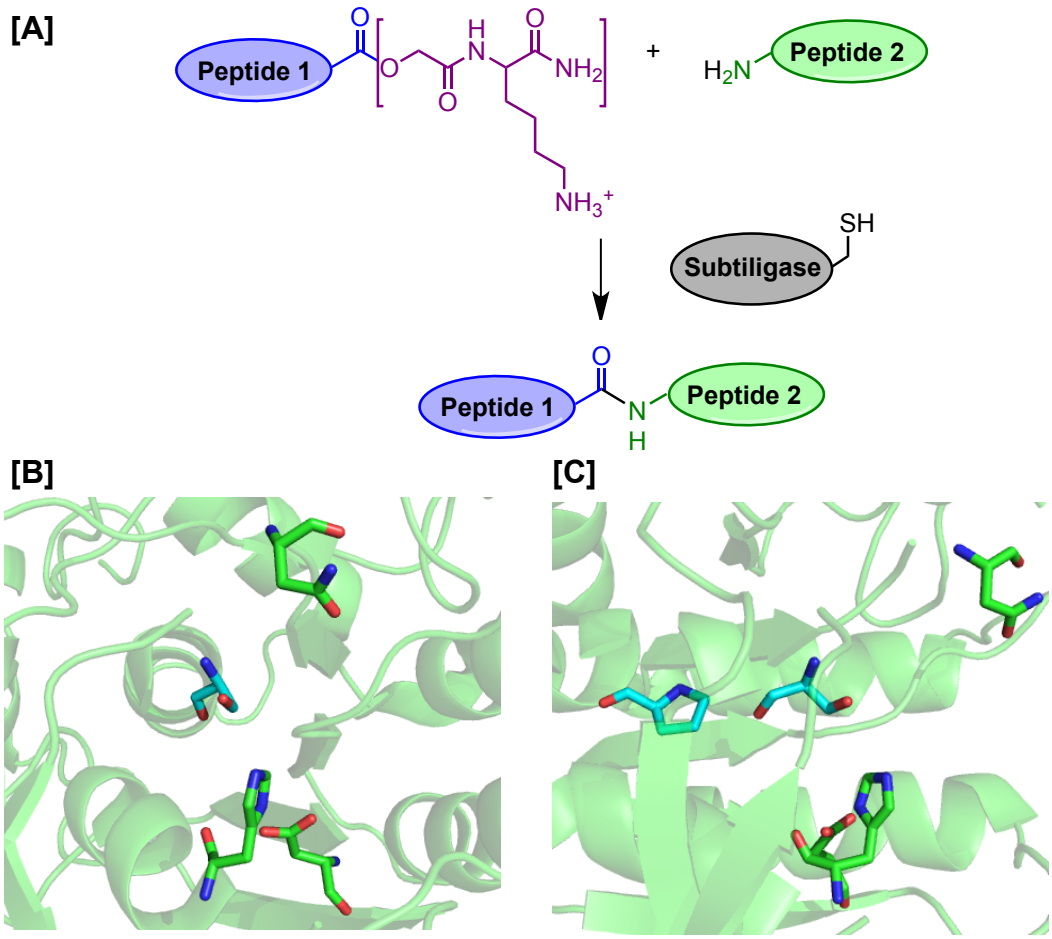


Figure 1-9. Subtiligase Reaction and Catalytic Triad in Wild Type Subtilisin
(**A**) The mutant subtilisin, subtiligase, prefers amide bond formation over lysis. (**B**) Catalytic triad is shown as sticks (S_{221} (in cyan), H_{64}, D_{32}) in WT subtilisin. (**C**) P_{225} is also shown in stick form in cyan to highlight that the mutation to Ala makes room in the enzyme active site.[16, 31] (PDB ID: 2SBT)[33]

1.2.7 – Non-amide Bond Forming Ligations and N-Terminal Modifications

Forming unnatural linkages provides options at the ligation site other than amide bonds (Figure 1-10). One ligation reaction that has been used *in vitro* and *in vivo* is click chemistry. Click reactions have a broad definition of being high yielding, selective, and kinetically favorable.[34] The standard click reaction developed by Sharpless and Meldal is a Cu^{I}-catalyzed reaction between an azido group and an alkyne, which forms a triazole ring.[35-36] The triazole is very stable and only 1.5 angstroms longer than an amide.[16] Click reactions can also be accelerated without the use of copper by using strained cyclooctynes developed by Carolyn Bertozzi (SPAAC reaction), which are more biocompatible reagents and have been used in live cell imaging.[37] Another option for peptide modification is the PLP-mediated biomimetic transamination, which is special for its chemoselectivity toward the N-terminus.[38] This reaction introduces a ketone or aldehyde at the α-amine of the N-terminus, and upon reacting with a substituted (R'-) hydroxylamine, forms an N-terminal oxime or hydrazine. The R' group of the hydroxylamine can effectively label the N-terminus with new functionality. This reaction can be done under reasonably mild aqueous conditions at pH 6.5. However, the reaction is incompatible with N-terminal serine, threonine, cysteine and tryptophan residues due to additional side reactions.[38]

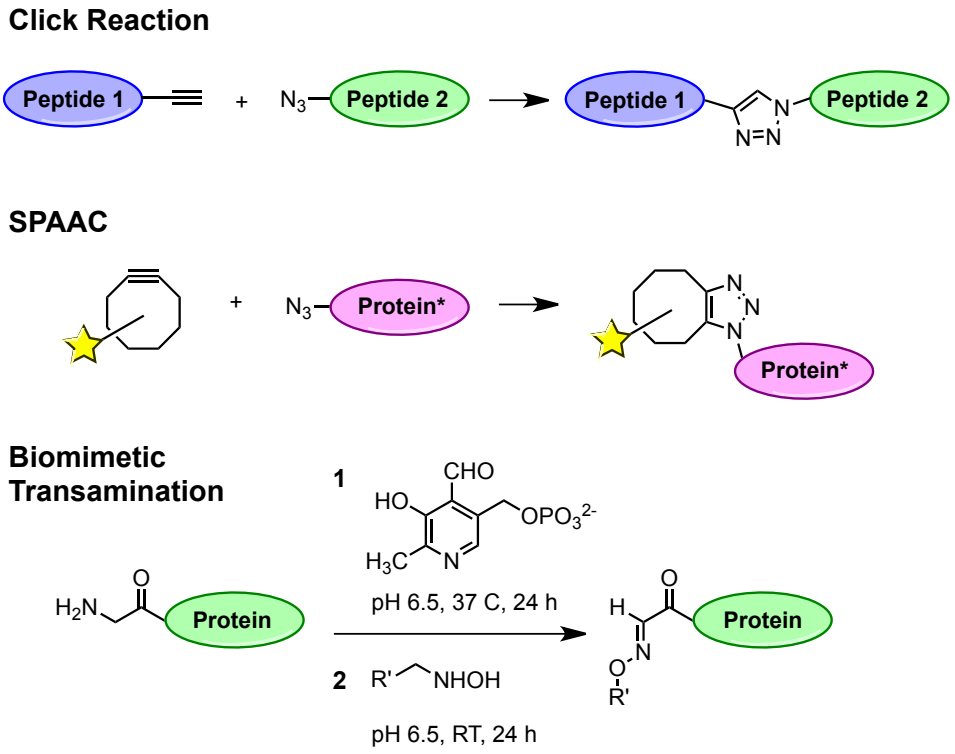


Figure 1-10. Alternate Non-Amide Bond Forming Reactions: Click Reaction, SPAAC, and Biomimetic Transamination

All of the aforementioned ligation and protein modification techniques have their respective strengths and weaknesses. There is a need for amide bond-forming reactions that allow for more flexibility at the ligation site, yet can still be easily performed at pH 7 in aqueous solution. To develop such an ideal reaction, we have investigated enzymatic catalysis by *E. coli* aminoacyl transferase (AaT). AaT catalyzes the formation of amide bonds using aminoacyl-tRNA as its substrate and transfers the amino acid onto positively-charged N-termini, i.e. Lys or Arg, of proteins independently of the ribosome.[39] In the following sections, I will discuss the biological basis for L/F transferase function and current research on the AaT active site.

Section 1.3 – AaT and the N-end Protein Degradation Pathway

1.3.1 – Prokaryotic N-End Pathway

AaT is the primary initiating enzyme in the N-end protein degradation pathway in *E. coli*.[40] The N-end pathway regulates protein half-life through the use of destabilizing amino acids, called degrons, to signal proteolytic machinery to degrade the "destabilized" protein.[40] This protein degradation pathway exists in both prokaryotes and eukaryotes. I will first discuss the role of AaT in the prokaryotic N-end pathway and then briefly discuss the similar pathway that exists in eukaryotes.

The purpose of protein degradation pathways in living organism is multifold: the regulation of short-lived signaling proteins and elimination of abnormal proteins that could be cell toxic.[41] The first degron (secondary destabilizing residue) in the N-end pathway in *E. coli* is lysine or arginine exposure of the N-terminus of a protein; these residues indicate metabolic instability.[42] An N-terminal lysine or arginine makes the protein a substrate for AaT, which is involved in the first step of the N-end pathway (Figure 1-11 and 1-12).[39] AaT specifically recognizes the positively charged N-terminal Lys or Arg with a negatively charged pocket within its active site.[43] AaT transfers leucine or phenylalanine (primary destabilizing residue) *via* a leucyl- or phenyalanyl-tRNA substrate to the protein N-terminus, making the second degron signal, N-terminal (L/F)(K/R).[42-43] An "adapter protein," ClpS, recognizes the second degron signal and brings the protein to the proteolysis machinery complex, ClpAP.[40, 44] According to Erbse *et al.*, ClpS has some affinity for the individual destabilizing residues, but ClpS has a higher affinity for the N-terminal sequence (L/F)(K/R), which correlated well with their *in vivo* degradation studies. Using a β-galatosidase-based assay, they found that F/L-R-

β-galatosidase was degraded faster than L-β-galatosidase.[44] The ClpS protein brings the two degron-tagged protein to the protease machine, ClpAP, a complex of two proteins ClpA and ClpP. ClpA is an AAA+ protein (ATPases associated with various cellular activities).[45] AAA+ proteins include a large group of proteins that have wide-ranging function, but have a connective structural element: an ATP-binding domain that is 200 to 250 amino acids in length.[45] AAA+ proteins couple the energy provided by their ATPase to an additional protein or macromolecule; in context of the ClpAP complex, ClpA is the "gate keeper" for proteins to access ClpP.[45-46] ClpS has an additional function in the *E. coli* N-end pathway; ClpS inhibits degradation of non-N-end rule proteins.[39] Upon binding of the degron-tagged protein, ClpA unfolds the protein *via* ATP hydrolysis, and passes it through a pore to ClpP, which processively degrades the protein into 7 to 10 residue peptide fragments without a significant degree of sequence specificity.[39, 47-49] Upon degradation, the peptide fragments can now be used in an amino acid "pool" for reuse in new protein synthesis.[41]

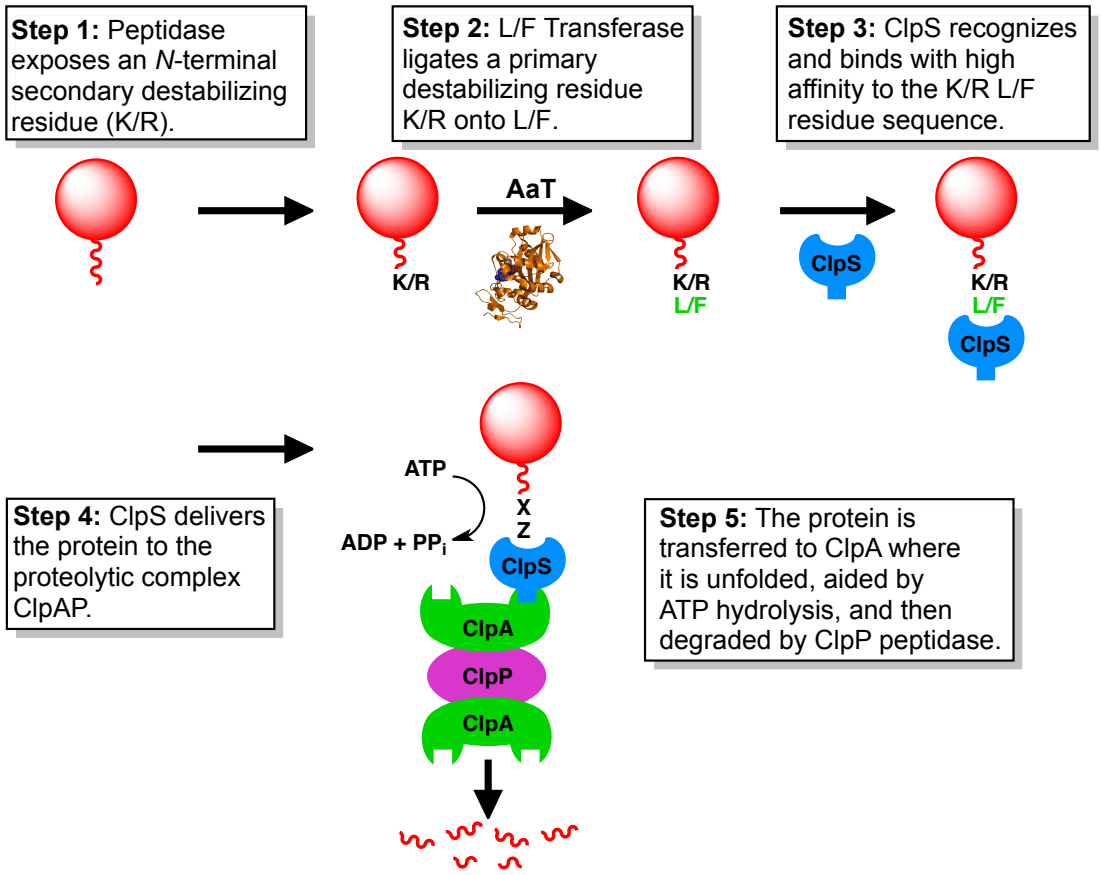


Figure 1-11. N-End pathway in E. coli
Adapted from Mogk *et al.*[39]

An additional transferase called Bpt that transfers only leucine has been identified in the N-end pathway of *Vibrio vulnificus*. *V. vulnificus* is prokaryotic pathogen found in marine environments, particularly shellfish.[50] Bpt has no sequence similarity to AaT but is a sequelog of Ate1, an arginyl transferase involved in the N-end protein degradation pathway in eukaryotes.[47] Bpt ligates leucine to the N-terminus of proteins bearing an N-terminal aspartate or glutamate (Figure 1-12). Additionally, *V. vulnificus* also has an AaT gene that is encoded in the genome directly adjacent to Bpt under the control of the same operon. Graciet *et al.* tested whether deletion of one of the transferase genes affected the

other gene's expression, and determined that each transferase could work independently.[50] According to coimmunoprecipitation tests and *E. coli* cAMP two hybrid analysis, Graciet *et al.* determined that it was unlikely that the two transferases interacted; however, the tandem gene of AaT-Bpt was found to be conserved within proteobacteria.[50-51] While key active site residues have been identified, the crystal structure has yet to be determined and more information is needed to understand the reason for transferase duplicity in *V. vulnificus*.

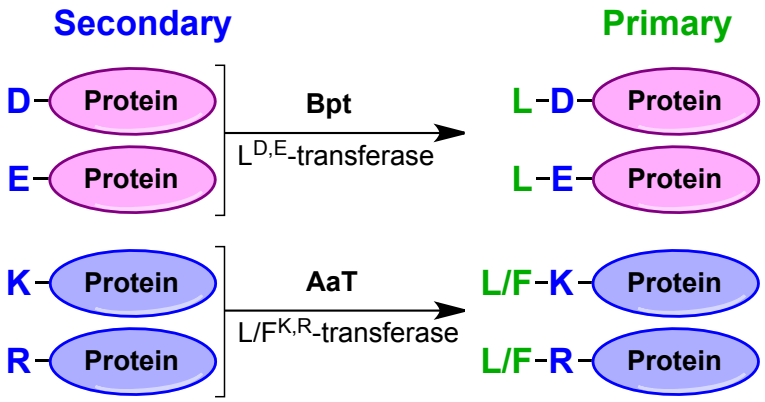


Figure 1-12. N-End Pathway N-degrons in Prokaryotes
Adapted from Varshavsky *et al*.[47]

1.3.2 – Eukaryotic N-End Pathway

Eukaryotes have a similar N-end protein degradation pathway but have additional N-degron signals (Figure 1-13).[52] In eukaryotes, N-terminal destabilizing residues are recognized by E3 ligases.[39] E3 ligases, with the help of E2 ligases and E1 ligases, allow for the polyubiquitilation of an internal lysine residue of the protein designated to be degraded by its N-terminal residue.[40] The multi-ubiquitilation acts as the adapter protein for eukaryotes and brings the protein to be degraded to the 26S proteasome (Figure 1-14).[39] The 26S proteasome acts similarly to the ClpAP machinery and is composed of

two complexes, 19S and 20S.[47] The 19S complex has 6 AAA+ proteins that are involved in the process of unfolding and passing the protein substrate to the 20S complex, which is the proteolytic core.[39] The origins of the ubiquitilation systems in the N-end pathway are still unclear, but the processive proteolysis found both in prokaryotes and eukaryotes likely links them by a common ancestor.[50]

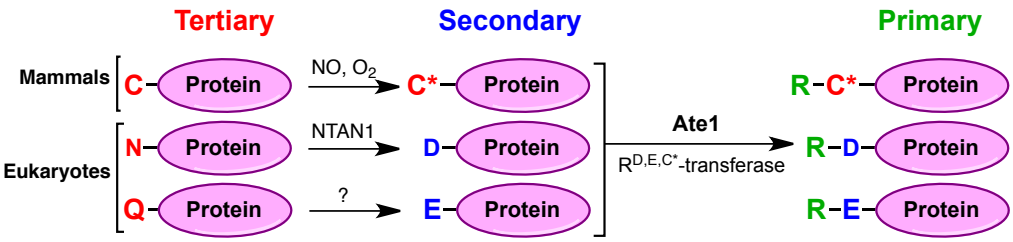


Figure 1-13. N-End Pathway N-degrons in Eukaryotes
Adapted from Varshavsky *et al*.[47]

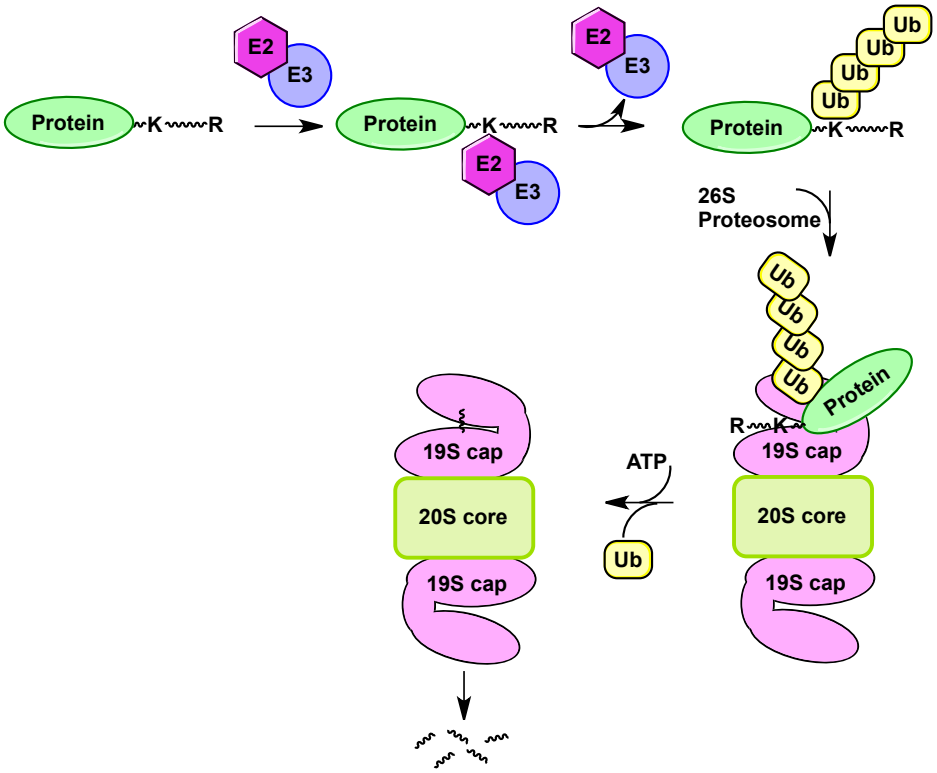


Figure 1-14. N-End Pathway in Eukaryotes
Adapted from Varshavsky *et al*.[47]

Section 1.4 – History of AaT and Its Active Site Chemistry

1.4.1 – History of AaT

The natural formation of amino acid peptide bonds can occur through a variety of mechanisms in both eukaryotes and prokaryotes with two well-studied processes being ribosomal protein synthesis and protein modification *via* aminoacyl-tRNA transferases. The first observation of transferase activity, discovered by Kaji *et al*. in 1965, was found thorough incorporation of radiolabeled leucine and phenylalanine in cell extracts of *E. coli*.[53] Five years later, the enzyme responsible for the amino acid transfer, AaT, was first extracted from *E. coli* and purified by Leibowitz and Soffer.[54] The unique characteristic of *E. coli* AaT was its ability to transfer amino acids onto proteins with equal preference for both leucine and phenylalanine.[54] They also determined several requirements of AaT function: optimal activity at pH 8, requirement of a monovalent cation source, 0.15 M KCl to achieve maximal kinetic function, enzymatic inhibition by divalent cations and the antibiotic puromycin, and complete independence of the ribosome, magnesium ions, and GTP.[54] In 1973, Soffer *et al*. found the protein acceptors of the AaT reaction to be N-terminal lysine or arginine, with stereoselectivity for the L-isomer and some loss of peptide substrate binding according to the penultimate N-terminal residue, in particular, glutamate.[55] In 1991, Tobias *et al*. determined that the physiological role of AaT was its involvement in the N-end rule in bacteria, but a crystal structure was yet to be determined until 2006 by Suto *et al*.[42, 56] According to the crystal structure at 2.4 angstrom resolution, AaT forms two distinct domains, an N-terminal domain and a C-terminal domain (Figure 1-15).[56] The crystal structure was solved with puromycin, a known inhibitor of AaT, bound in its active site, which is a cleft formed

between the N- and C-terminal domains.[56] The structural fold of the C-terminal domain is similar to enzymes in the FemABX family. The FemABX family of enzymes contain a GNAT fold and their active sites catalyze the binding of an amino acid, using aminoacyl tRNA, to a peptidoglycan precursor, which is involved in cell wall development.[56-57] Despite the similarity in aminoacylation reactions, the FemABX enzymes hold no significant primary sequence similarities to AaT.[56]

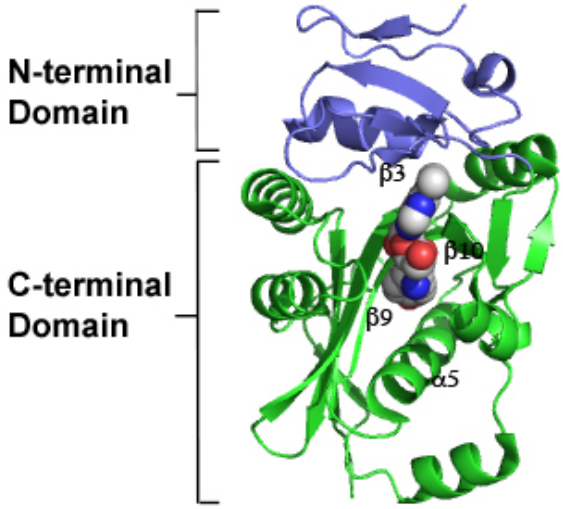


Figure 1-15. Crystal Structure of AaT at 2.4 Å
The N-terminal domain consists of amino acids 3-62 (four β-strands, β1–4, and one α-helix, α1) and the C-terminal domain consists of amino acids 63-232 eight β-strands, β5–12, and six α-helices, α2–7). The chains that make up the active site of AaT are from both the N and C terminal domains (β9, α5, and β10 of the C-terminal domain and β3 from the N-terminal domain). PDB ID – 2DPT.[56]

1.4.2 – AaT Active Site Chemistry

AaT naturally transfers three amino acids, leucine and phenylalanine, which are involved in the N-end pathway, and methionine at a lower efficiency.[58] Using the crystal structure and amino acid mutagenesis, amino acid residues in the binding cleft, M_{144}, F_{153}, L_{170}, F_{173}, and I_{185}, were found to interact with the hydrophobic portion of

puromycin, which could indicate the leucyl/phenylalanyl binding site of the aminoacyl tRNA.[56] Suto *et al.* proposed that the active site discriminates against β-branched residues such as isoleucine due to the close proximity (≤ 4 Å) of Met_{158} and the amide bond of Gly_{155} and Glu_{156} (Figure 1-16). Additionally, they proposed a tRNA docking model with AaT proposing that the tRNA 3' end likely adopts a bent conformation that mediates amino acid binding in the hydrophobic pocket of AaT.[56]

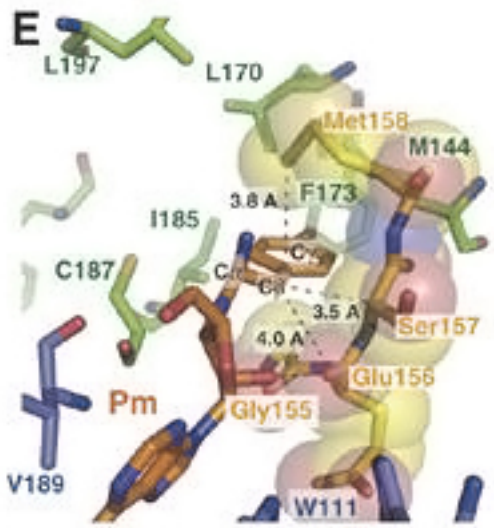


Figure 1-16. Amino Acids Involved in Puromycin Binding in AaT
The aminoacyl moiety (C_α, C_β, C_γ) is stacked between residues Met_{158} and the peptide bond linking Gly_{155} and Glu_{156}. The Hydrophobic pocket: Met_{144}, Phe_{153}, Leu_{170}, Phe_{173}, Ile_{185}. Pm = puromycin. Reprinted with permission from (Kyoko Suto, Yoshihiro Shimizu, Kazunori Watanabe, Takuya Ueda, Shuya Fukai, Osamu Nureki, Kozo Tomita. *The EMBO Journal.* **2006**, Vol. 25, No. 24, 15139–15147.). Copyright (2006) Nature Publishing Group.

More information about the reaction mechanism was elucidated in 2007 by the same group using crystal structures bound separately to minimized AaT substrates, phenylalanyl adenosine and peptide product, FRYLG. Watanabe *et al*. determined that Q_{188}, acting as a general base, follows a mechanism similar to the reverse reaction of a serine protease (Figure 1-17).[43] However, the mechanism was challenged by Fung *et al*.

in 2011, and through the use of highly sensitive MALDI analysis, it was determined that none of the active site residues are actually catalytic; instead, the active site simply allows for "substrate-assisted catalysis" of the substrates, in which the α-amine of the protein substrate undergoes nucleophilic attack on the ester carbonyl bond of the aminoacyl L-tRNA^{Leu} or Phe-tRNA^{Phe}. Currently, the proposed mechanism by Fung *et al*. seems more likely to be correct because the supposed catalytic residue of the Watanabe *et al*. proposed mechanism, Q_{188}, can be mutated to an alanine and still retain enzymatic activity; however, neither mechanism has been fully confirmed or disproved.[43, 59]

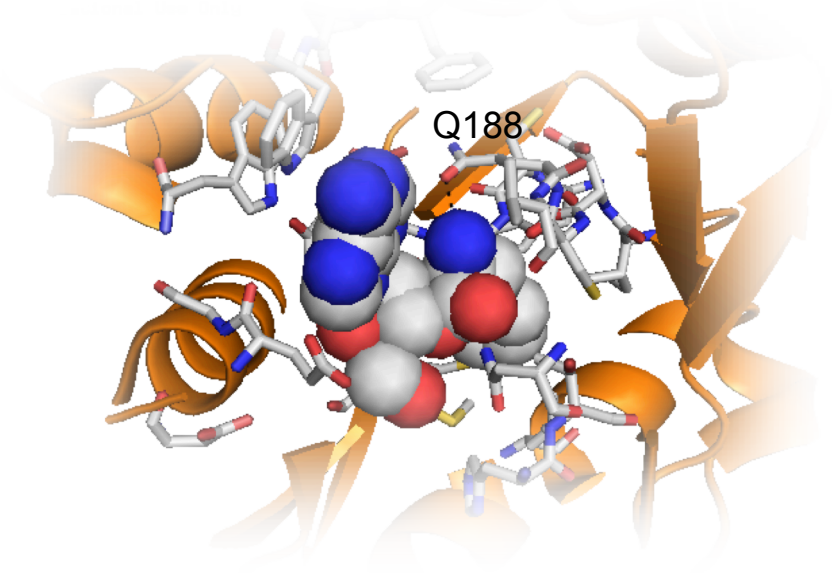


Figure 1-17. Phe-A Bound to the Active Site of AaT
Q_{188} is labeled to show its close proximity to the Phe-A substrate, 2.9 Å. PDB – 2Z3K[43]

Section 1.5 – Minimization of AaT Substrates

The selection of the correct amino acids in ribosomal protein synthesis is complex, with a variety of checks and balances throughout the process.[60] Gromadski *et al*. determined that the ribosome exhibited a greater than 450 fold selectivity for cognate aminoacyl tRNA anticodons versus a single mismatch at the anticodon.[60] Effraim *et al*. tested the ribosomal acceptance of combinations of Ala, Phe, and Lys tRNAs and found that misacylation does not effect the yield of dipeptide formation when only misacylated tRNA is supplied; however, when in competition with correctly acylated tRNA, the ribosome shows a clear discrimination against misacylation.[61] Similarly, AaT has been shown to use misacylated tRNA as well as minimized tRNA mimics (Figure 1-18).[62-63] The ability of AaT to use misacylated non-cognate tRNA indicated that amino acid identity was less selective than previously expected.[64] Abramochkin *et al*. determined that the aminoacyl-tRNA anticodon is not a transferase recognition determinant.[64] Also, mutations in the anticodon and extra arm of the $tRNA^{Leu}$ substrate did not substantially alter transfer.[64] They did not find stringent conservation of nucleotide sequences between L/F transferase tRNAs, but charged tRNA with a GC stem was a poor substrate.[64] Connor *et al*. used mutant aminoacyl tRNA synthetases to transfer unnatural amino acids to the N-termini of peptides using AaT.[65] Taki *et al.* also minimized the aminoacyl tRNA substrate and demonstrated that truncated aminoacyl tRNAs from 20 bases down to dinucleotides could be utilized as an AaT substrates.[63] The greatest minimization of the AaT aminoacyl tRNA substrate was reduction to phenylalanyl adenosine (Phe-A), which was used as a substrate analog for analysis of substrate binding in crystal structures.[43]

Watanabe *et al.* found that in the presence of an accepter peptide AaT could use Phe-A as a substrate, forming approximately 1 % N-terminally phenylalanine-ligated product.[43] These findings have shown that AaT can have amino acid promiscuity in its active site while maintaining ligation efficiency for a variety of tRNA lengths.

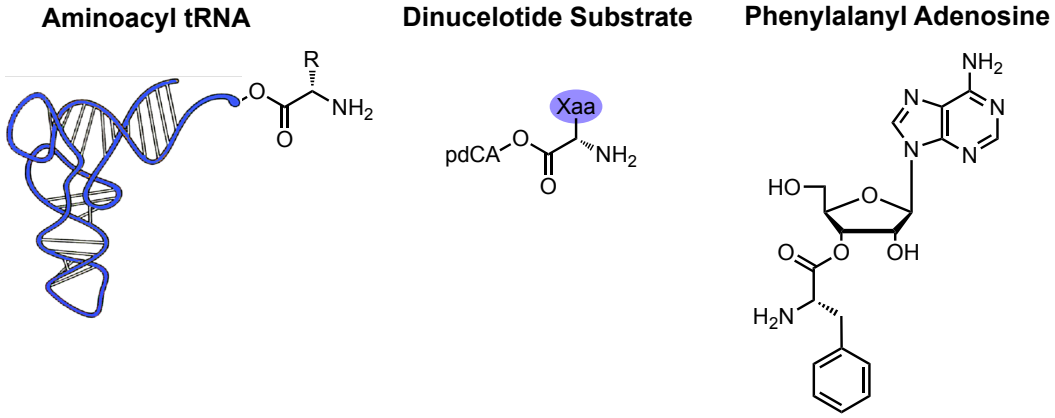


Figure 1-18. Minimized Substrates
Aminoacyl tRNA, specifically L/F, is the natural substrate for AaT. Dinucleotides (Xaa = amino acid) and phenylalanyl adenosine have also been shown to be AaT substrates.

Section 1.6 – Summary

Protein modification is an important tool for understanding how a protein moves, folds, and operates in its cellular environment. However, many current N-terminal modifications and ligation techniques can perturb protein structure from its natural state through cysteine mutagenesis and unnatural amide backbone linkages. In our work, we aim to learn more about the protein structure and function with minimal perturbation from its natural state. Our new method of protein modification ligates small molecules onto proteins with N-terminal specificity and creates natural amide bonds at ligation sites. AaT-mediated ligations recognize a small synthetic aminoacyl-tRNA analog, aminoacyl adenosine, which can be chemically synthesized in only three steps.[66] Using AaT, we

can enzymatically attach a wide range of synthetic moieties to the N-termini of peptides and proteins, including natural and unnatural amino acids with reactive azido groups, fluorescent labels, protected disulfides, and photoinducible crosslinkers. This method introduces new opportunities for protein semi-synthesis and contributes an effective method to label proteins for biophysical research, which we have demonstrated in two model proteins, α-synuclein and calmodulin. We have also explored other aminoacyl transferase options, Bpt from *V. vulnificus* and Ate1, a mammalian arginyl transferase.

CHAPTER 2

N-TERMINAL PROTEIN MODIFICATION USING SIMPLE AMINOACYL TRANSFERASE SUBSTRATES

Section 2.1 - Introduction

The conjugation of synthetic molecules to proteins contributes to biomedical research by enabling the immobilization of proteins on surfaces, modifications with chromophores to make *in vitro* sensors, or tagging with *in vivo* imaging agents.[67-71] In most cases, the formation of well-defined conjugates is valuable, if not essential. The protein termini are attractive targets for conjugation, because in many cases appending synthetic molecules at the termini will have minimal undesired effects on protein folding and function.[72-92] This is particularly applicable in cases where the terminus is relatively unstructured. As part of our research on minimalist labeling strategies for biophysical studies, we sought a method for N-terminal modification that could be carried out easily under conditions that maintain folding without substantial prior protein manipulation.[12]

Selective N-terminal modification has been achieved by a variety of chemical and enzymatic methods, each with benefits and drawbacks. Small molecule strategies permit the attachment of a variety of molecules, but they are subject to side reactions and incomplete specificity for the N-terminus and may need to be carried out in organic solvent mixtures.[38, 82, 93-101] Reverse proteolysis methods can be used to modify the N-terminus under conditions that do not require protein unfolding, but the reaction can be difficult to drive to completion without high protein concentrations.[23, 32, 102-113] Several

other chemoenzymatic methods are valuable in functioning under benign conditions but have moderate-sized target sequences that must be appended to the protein.[114-121] Here, we describe a minimal system for N-terminal protein labeling that utilizes only adenosine esters of natural or unnatural amino acids and a single, readily available enzyme. This process gives high yields of modified proteins under nondenaturing conditions and requires only a single basic amino acid for specific recognition.

Aminoacyl tRNA transferases (AaTs) are members of a growing class of enzymes that use aminoacyl tRNAs in secondary metabolism.[122] The *E. coli* AaT catalyzes the transfer of Leu, Phe, or Met from an aminoacyl tRNA to a protein bearing an N-terminal Arg or Lys.[43, 56, 123] The addition of Leu or Phe targets that protein for degradation by ClpA as part of the N-end rule pathway.[39-40, 42, 124] Kaji *et al.* first observed AaT aminoacylation activity in crude *E. coli* preparations.[53, 123, 125] Soffer and Leibowitz subsequently reconstituted the purified enzyme and characterized its specificity or both the RNA and amino acid components of the donor molecule, demonstrating its first use in transferring an unnatural amino acid, *p*-fluorophenylalanine (Figure 2-1, Top).[126-127]

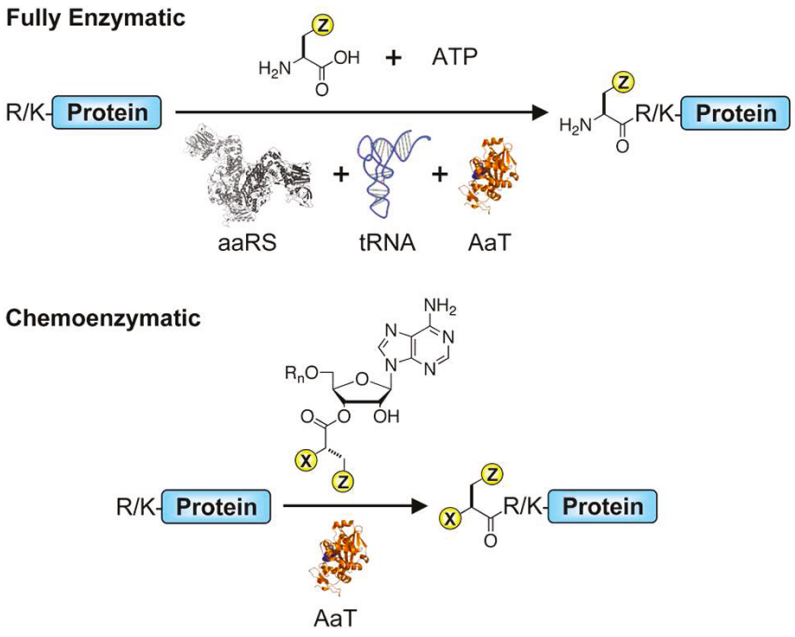


Figure 2-1. Transferase-Mediated N-terminal Protein Modification
(Top) Fully enzymatic methods use aminoacyl tRNA synthetase (aaRS), tRNA, aminoacyl transferase (AaT), amino acid, and ATP. (Bottom) Chemoenzymatic methods use only a synthetic nucleic acid donor and transferase. In this work, an aminoacyl mononucleoside (R_n =H) was used.

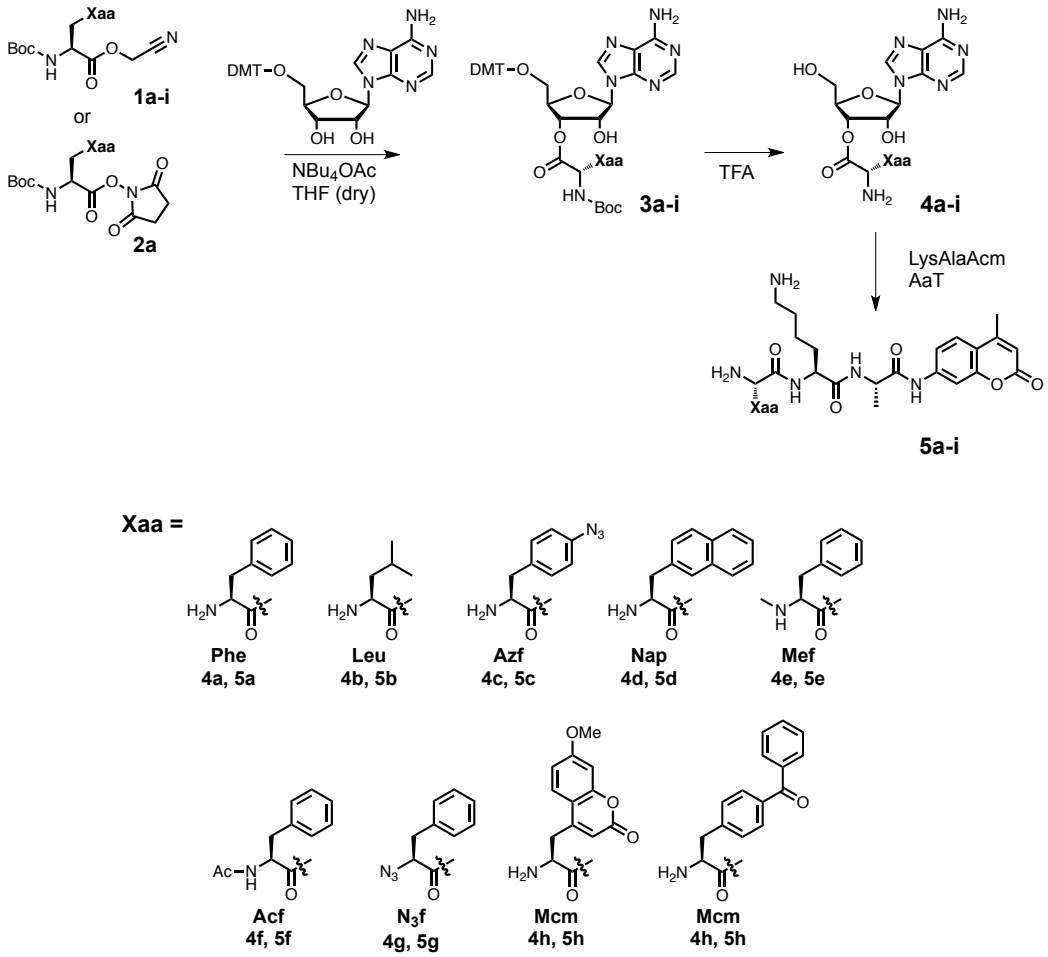
Recently, purified AaT has been used to modify proteins *in vitro* with a variety of unnatural amino acids charged onto tRNAs, either by chemical semi-synthesis or the use of a mutant aminoacyl tRNA synthetase (aaRS).[62-63, 85, 128-129] Abramochkin and Shrader discovered that AaT was tolerant of variation in the acceptor stem (aminoacylation site) of the tRNA, and Sisido and coworkers have shown that much shorter oligonucleotides (Figure 2-1, bottom, R_n = 1 - 21 nucleotides) can act as donors.[63-64, 130] As part of their structural characterization of AaT, Tomita and co-workers found that AaT could bind phenylalanyl adenosine and transfer Phe to peptides in trace amounts.[43] However, the use of adenosine mononucleoside as a donor substrate for AaT has not, to our knowledge, been further explored.

We have found that a variety of hydrophobic amino acids can be transferred from adenosine donors with high yields at relatively low protein concentrations (Figure 2-1, bottom, R_n = H). These donors can be synthesized from commercially available active esters of the amino acids. Reduction of the two step synthetase/AaT reaction sequence to the one step AaT-only sequence allows us to explore the substrate specificity of AaT without limitation by the specificity of the synthetase in the first step. While this is also possible with Sisido's pdCpA dinucleotide donors, the shorter synthesis of the simple adenosine donors (1 - 2 steps *vs.* 7 steps) makes this exploration easier. We have characterized the substrate scope of the wild type AaT enzyme using a reporter peptide, determined the kinetic parameters for the adenosyl substrates, and demonstrated the efficient modification of a full-sized protein, 23 kDa α-casein.

Section 2.2 - Results and Discussion

The synthesis of adenosyl donor compounds, shown in Scheme 2-1, can be carried out on any amino acid with appropriate acid-labile protecting groups (e.g., *N*-Boc). This synthesis is not limited to α-amino acids, and our donor synthesis has also enabled us to test the transfer of analogs such as α-azidophenylalanine (N_3f, 4g). Conversion to the cyanomethyl ester (1a - h) is typically high-yielding, although yields for the subsequent acylation of 5′-*O*-dimethoxytrityl-adenosine ((DMT)-A) varied between 60 and 90 %. Acylation yields generally decreased with side chain bulk; derivatives such as N-1,8-naphthyldiaminopropionic acid coupled poorly (data not shown). Uncatalyzed reactions took 1 - 5 days, but in all cases, addition of

tetrabutylammonium acetate increased reaction rates.



Scheme 2-1. Synthesis of Adenosyl Donors and Use in AaT-Catalyzed Protein Modifications

Donor synthetic intermediates: 1a, 2a, 3a X = NHBoc, Z = Ph; 1b, 3b X = NHBoc, Z = i-Pr; 1c, 3c X = NHBoc, Z = *p*-azidophenyl; 1d, 3d X = NHBoc, Z = 2-naphthyl; 1e, 3e X = NHMeBoc, Z = Ph; 1f, 3f X = NHAc, Z = Ph; 1 g, 3 g X=N_3, Z = Ph; 1 h, 3 h X = NHBoc, Z = 7-Methoxycoumarinyl; 1i, 3i X = NHMeBoc, Z = *p*-benzoylphenyl

The Phe donor **4a** was synthesized from both the cyanomethyl ester (**1a**) and *N*-hydroxysuccinimidyl (OSu) ester (**2a**). All other adenosyl donors were synthesized from the cyanomethyl ester (**1b - h**). After trifluoroacetic acid (TFA) deprotection and CH_2Cl_2

washing (or Et_2O precipitation) to remove DMT byproducts, the compounds could be taken on directly to AaT enzymatic reactions (contaminants were limited to incompletely deprotected compounds which are not transferred by AaT). However, in most cases the deprotected adenosyl donor was purified by HPLC, and higher transfer yields were generally observed with these purified donors.

Transfer extent was evaluated by HPLC analysis of reactions with adenosyl donor, AaT, and an aminocoumarin-labeled reporter peptide used previously by Tirrell and co-workers (LysAlaAcm, Scheme 2-1).[65] Monitoring LysAlaAcm acylation in crude reaction mixtures using absorbance at 325 nm conveniently eliminates background due to other proteins and nucleic acids, although some unnatural amino acids such as napthylalanine (Nap) and methoxycoumarinylalanine (Mcm) also absorb in this range. Example HPLC chromatograms from Phe ligations are shown in Figure 2-2. LysAlaAcm starting material eluted at 11.8 min (Figure 2-2A); hydrophobic products eluted at 12 to 15 min (PheLysAlaAcm **5a** elutes at 13.0 min in Figure 2-2B, NapLysAlaAcm **5d** elutes at 14.1 min in Figure 2-2D).

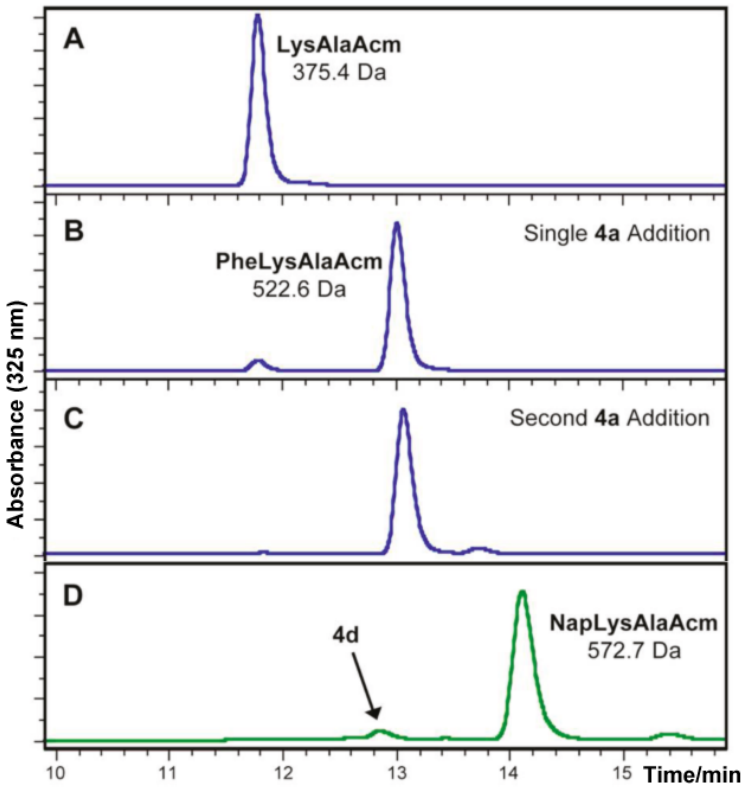


Figure 2-2. Reversed-Phase HPLC Analysis of Transferase Reactions
Top: AaT-mediated transfer of Phe from adenosyl donor **4a** to LysAlaAcm reporter peptide. HPLC chromatograms obtained after (A) 0 h or (B) 4 h show conversion of LysAlaAcm (11.8 min retention time) to PheLysAlaAcm (13.0 min retention time). Conversion after one addition of donor **4a** is 92 % (B). A second addition of 1 mM **4a** drives the reaction to completion with a small amount of double Phe addition (C). Bottom: HPLC analysis of transfer of Nap from adenosyl donor **4d** to LysAlaAcm after (D) 4 h shows conversion to NapLysAlaAcm (14.1 min retention time). Product identities confirmed by observation of indicated masses by MALDI MS.

To evaluate substrate scope, reactions were analyzed after 4 h, when transfer should be complete for the fully enzymatic reaction with Phe tRNA synthetase.[65] Phe (**4a**) and Leu (**4b**) were transferred efficiently (Table 2-1). Our PheLysAlaAcm yields from donor 4a were comparable to yields from Phe-pdCpA, the dinucleotide used by Sisido. Transfer yields for other substrates varied with side chain size, where Mcm (**4h**) is too large and benzoylphenylalanine (Bzf, **4i**) is a poor substrate (see Table 2-1). Sisido has shown that AaT mutants can use substrates with larger side chains; we are currently

exploring some of these mutations.[63] While transfer yields after 4 h varied, most reactions could be driven to completion by a bolus of donor molecule (Figure 2-2C).

Table 2-1. N-Terminal Modification of Reporter Peptide After Single Addition of Adenosyl Donor

Donor	Additives[a]	Yield (%)[b]
4a (Phe)		92.7 ± 5.2
Phe-pdCpA		95.2 ± 4.3
4a (Phe)	1 mM ATP	86.1 ± 0.8
4a (Phe)	5 mM ATP	79.1 ± 0.4
4a (Phe)	5 mM AMP	88.2 ± 0.8
4a (Phe)	5 mM A	93.1 ± 1.1
4a (Phe)	1 mM pdCpA	90.4 ± 0.9
4b (Leu)		80.4 ± 0.9
4c (Azf)		78.5 ± 5.2
4d (Nap)		95.4 ± 0.4
4d (Nap)	100 μM PheLysAlaAcm	93.8 ± 0.3
4e (Mef)		0.6 ± 0.1
4f (Acf)		5f not observed
4g (N_3f)		5g not observed[c]
4h (Mcm)		5h not observed
4i (Bzf)		9.2 ± 0.3

[a] Standard AaT-catalyzed transfer conditions with or without potential inhibitor added. [b] Yield determined by integration of HPLC chromatogram peak areas at 325 nm, described in Materials and Methods. Standard deviations are reported for an average of at least 6 experiments with at least 2 different AaT preparations. [c] HPLC and MALDI MS analysis indicate that the observed product is **5a** not **5g**.

One of the advantages of adenosyl donors over in situ tRNA aminoacylation with a tRNA synthetase is that we are not limited by the substrate specificity of the synthetase. In particular, nonamino acid substrates can be attached to adenosyl donors to assess their transferability by AaT. *N*-Methyl phenylalanine (Mef, **4e**) can be transferred, albeit with

low yield. We have also tested *N*-acetyl phenylalanine (Acf, **4f**) and α-azidophenylalanine (**4g**) and found them to be poor substrates. In some experiments, substantial conversion of LysAlaAcm was found when incubated with **4g**, but MALDI MS analysis has shown the product to be **5a**, presumably obtained by transfer of Phe after prior reduction (i.e., Phe donor **4a** is formed from **4g** under the reaction conditions). Steps taken to alter the concentration of β-mercaptoethanol reductant did not improve the yield of **5g** (data not shown). We believe that disproportionation of the azide is incomplete, but that AaT is selective for the α-amine substrate (Phe-A, **4a**), so far more **5a** than **5g** is formed. This may be due to an important interaction of the α-amine with Q_{188}, which also acts as a catalytic base to deprotonate the peptide substrate for amide bond formation. The Q_{188}/α-amine interaction was defined by the structural and enzymological work of Watanabe and co-workers.[43] We note that the preference of AaT for unmodified N-termini was previously documented with full-length tRNA by Soffer and Leibowitz.[127]

To better understand limitations on the yield from AaT reactions, we assessed the possibilities of inhibition by hydrolyzed substrate or aminoacylated product. Since the acyl adenosine donor substrate is not regenerated during the reaction (unlike the aaRS/tRNA/AaT reaction), we hypothesized that the adenosine byproduct might inhibit further acylation by binding to AaT. To investigate this possibility, we added adenosine to the reaction at varying concentrations and monitored the reaction timecourse by fluorescence spectroscopy and by HPLC. Addition from the Phe donor **4a** was not inhibited by the addition of 5 mM concentrations of adenosine (Table 2-1). A

comparison of adenosine compounds, including pdCpA, the dinucleotide donor used by Sisido, showed that only adenosine triphosphate (ATP) inhibited the reaction.

We were also able to monitor product formation in real time based on partial quenching of coumarin fluorescence upon addition of Nap to the LysAlaAcm peptide to form **5d**. The fluorescence intensity at 390 nm of NapLysAlaAcm is 28 % of the fluorescence of LysAlaAcm, and the overall fluorescence of mixtures of the two peptides can be used to determine the proportions of each peptide if the total concentration is known. Intensity was used to monitor NapLysAlaAcm formation in real time, and HPLC injection of the reaction end points were used to confirm the final product distribution. We found that exogenous adenosine did not significantly inhibit the reaction, even at concentrations up to 1 mM (Figure 2-3A).

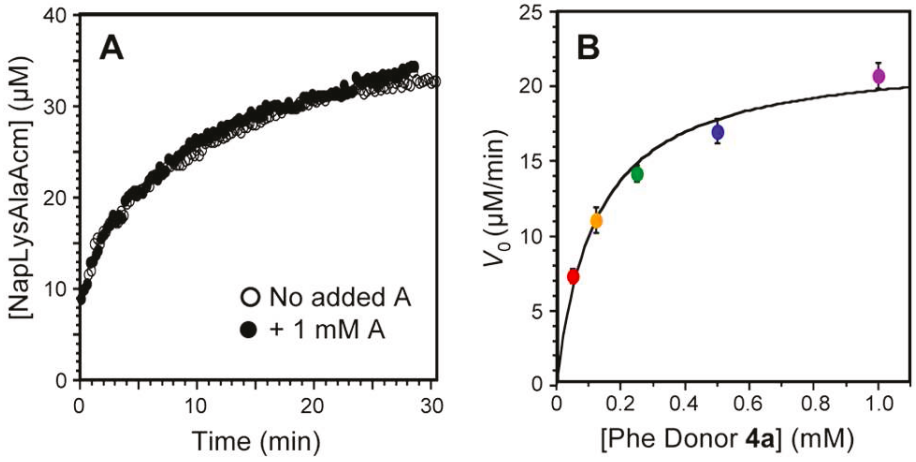


Figure 2-3. Kinetic Analysis of AaT Reactions
(A) Real time monitoring of Nap transfer to LysAlaAcm from 4d by quenching of Acm fluorescence. No significant inhibition is observed in the presence of 1 mM adenosine. (B) Saturation curve used to determine Michaelis-Menten kinetic parameters for LysAlaAcm modification by Phe donor **4a**.

We also assessed the possibility of product inhibition by the aminoacylated

LysAlaAcm product. In our standard 4 h transfer assay, addition of 100 μM (1 equiv.) PheLysAlaAcm did not significantly inhibit production of NapLysAlaAcm (**5d**) when Nap-A (**4d**) was used as a donor (Table 2-1). MALDI MS analysis and differences in HPLC retention time allowed us to easily distinguish LysAlaAcm and PheLysAlaAcm from NapLysAlaAcm (Figure 2-2). Finally, we investigated the question of whether donor hydrolysis slowed reaction rates after 30 - 60 min. Analysis of mock reactions lacking transferase or LysAlaAcm indicates that donor hydrolysis may contribute significantly to slowing reactions, which is why a donor bolus can be used to drive the reaction to completion (see Materials and Methods section).

To characterize the enzymatic activity of AaT toward our acyl adenosyl substrates, we used our HPLC assay to determine initial reaction rates to be fit to a standard Michaelis-Menten kinetic model. Equations describing this kinetic analysis are included in the Supporting Information. Since adenosine does not inhibit the enzyme, by keeping LysAlaAcm concentrations high (100 μM), we measured the rate of transfer as a function of the concentration of Phe donor **4a** and fit this to a single-substrate kinetic model to determine a k_{cat} of $1.68 \pm 0.09 \times 10^{-1}$ s^{-1} and a K_M of $1.24 \pm 0.23 \times 10^{-4}$ M. The enzyme efficiency ($k_{cat}/K_M = 1.35 \pm 10^3$ M^{-1} s^{-1}) is relatively low.[131] While K_M is higher for Phe-A than what Shrader reported for aminoacyl $tRNA^{Leu-4}$ (0.3 μM), the k_{cat} is comparable (0.13×10^{-1} s^{-1}).[64] Since it is trivial to run our reactions at donor concentrations well above K_M, the lower apparent affinity for our substrates should not affect the utility of the reaction. We note that the binding affinity, which gives a K_M of 124 μM for **4a** must arise primarily from Phe binding interactions since the K_I for

adenosine inhibition is greater than 5 mM.

It is possible that larger protein substrates are modified with different efficiencies than peptide substrates.[63-64, 127] To address this question and to demonstrate the utility of our method in tagging full-sized proteins with useful fluorophores or affinity purification tags, we modified the milk protein α-casein (bearing a native N-terminal Arg after proteolysis of a leader sequence).[132] First, modification using Phe donor **4a** was assessed by Edman degradation, where successful, quantitative addition of Phe was observed.

In a more complex set of experiments, we first modified α-casein with Azf using donor **4c** and then used a Cu-catalyzed Huisgen cycloaddition ("click" reaction) to append a fluorescein label (Figure 2-4A).[133-136] Prior to fluorescent-labeling, the initial transfer of Azf was analyzed by trypsin digest and MALDI MS of α-casein (Figure 2-4C). The N-terminal peptide ArgProLys disappeared from the mass spectrum following modification and a peptide corresponding to AzfArgProLys was observed (also a mass corresponding to AzfArgProLys after loss of N_2 during ionization; N_2 loss was also observed for AzfLysAlaAcm). Complete conversion of α-casein was observed in MALDI MS due to double addition of **4c** to drive the reaction.

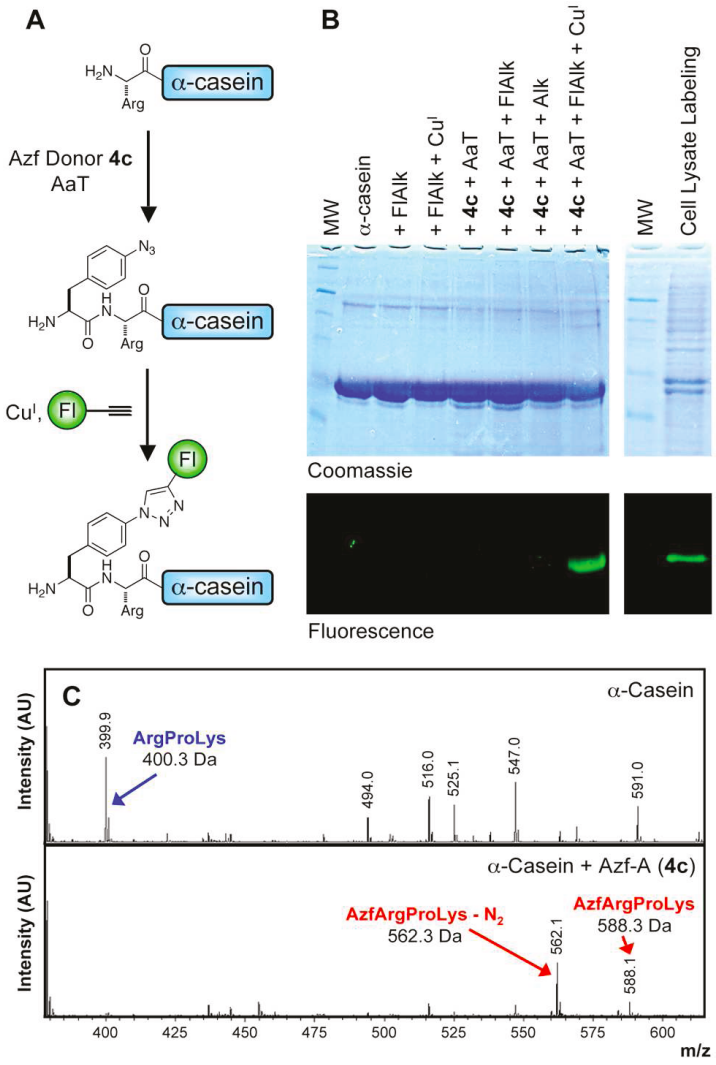


Figure 2-4. AaT-Catalyzed Modification and "Click" Reaction of α-Casein N-terminus
(A) α-Casein modification scheme. (B) PAGE gel analysis of α-casein modification. Lanes (left to right): (1) molecular weight (MW) markers (Masses in kDa: 17, 25, 30, 46, 58, 80, 175); (2) α-casein; (3) α-casein mixed with fluorescein alkyne (FlAlk); (4) α-casein mixed with FlAlk, $CuSO_4$, THPTA, and sodium ascorbate; (5) α- casein mixed with Azf-A (**4c**) and AaT; (6) α-casein mixed with 4c and AaT, then FlAlk; (7) α-casein mixed with 4c and AaT, then propargylamine (Alk), $CuSO_4$, Tris-(3-hydroxypropyltriazolylmethyl) amine (THPTA), and sodium ascorbate; (8) α-casein mixed with **4c** and AaT, then FlAlk, $CuSO_4$, THPTA, and sodium ascorbate;(9) MW markers; and (10) cell lysate labeling, α-casein reaction carried out using conditions of lane 8 with unpurifed AaT in cleared *E. coli* lysate. (C) MALDI MS analysis of trypsinized N-terminal fragment of α-casein with or without modification by Azf using AaT and **4c** (double addition).

After AaT-catalyzed modification of α-casein with Azf, residual **4c** was removed by dialysis, AaT (His_{10}-tagged) was removed by treatment with Ni beads, and the azide-

bearing protein was reacted with fluoroescein alkyne (FlAlk) in the presence of Cu(I). The results of this labeling experiment, as well as appropriate control reactions, are shown in Figure 2-4B. One can see that fluorescent α-casein is only observed under the proper conditions (Lane 8), where both the chemoenzymatic labeling by AaT with **4c** and the subsequent click reaction are successful. Finally, we carried out a similar labeling experiment in crude *E. coli* lysate with unpurified AaT. Azf transfer was successful, as shown by the observation of fluorescence after labeling with FlAlk. Control reactions (shown in Supporting Information) show no observable labeling of any protein, implying that endogenous levels of proteins terminating in Arg or Lys are too low to be observed here.

Substantial recent effort has been devoted to developing methods for labeling proteins under conditions that maintain protein folding and activity.[16, 137-138] *E. coli* aminoacyl tRNA transferase can work under these conditions and specifically modifies the N-terminus, a useful point for conjugation. However, certain aspects of its native mechanism are nonideal for its use as a preparative tool (i.e., a multistep, multienzyme reaction sequence). The work of Leibowitz, Shrader, and Sisido showed that *E. coli* aminoacyl tRNA transferase can modify the N-terminus of a protein ending in Arg or Lys with amino acids from an aminoacyl oligonucleotide donor.[63-64, 127] Watanabe *et al.* crystallized AaT with Phe-A bound and showed that Phe could be transferred from Phe-A to the N-terminus of the peptide RYLGYL in trace amounts. These two results provide crucial precedent for our own work, in which we show that AaT can efficiently use a variety of easily synthesized acyl adenosine substrates. Enzymological analysis

demonstrates that AaT is able to use the Xaa-A substrates with turnover numbers comparable to full-length aminoacyl tRNAs. Furthermore, we have shown that neither the adenosine reaction byproduct nor aminoacylated protein product (at least in the case of our reporter peptide) appreciably inhibit the reaction. In fact, the major obstacle to efficient transfer is donor hydrolysis, but this can be overcome by a second addition of donor molecule since the K_I for inhibition by adenosine is greater than 5 mM.

Our AaT-only reaction sequence allows us to explore the compatibility of Phe analogs lacking a primary α-amine with transfer by AaT. The wild type transferase seems to have a strong preference for the α-amine, as only *N*-methylphenylalanine was transferred. In contrast, the side chain pocket is quite permissive: we have observed the transfer of bicyclic amino acid side chains and Sisido has reported mutants capable of transferring tricyclic amino acids.[63] We are currently investigating the transfer of larger side chains such as benzophenone, biotin, and fluorescein derivatives.

Our reaction conditions and yields compare favorably to two well-established methods in the field. Subtiligase (reverse proteolysis) reactions can be carried out in a few hours, but rarely go to completion.[108-109, 113] The pryidoxyl-5-phosphate (PLP) transamination reactions of Francis and co-workers can require elevated temperatures to achieve high yields and are subject to a number of side reactions.[100-101] AaT transfer reactions are highly specific (no Lys side chain modifications were observed in our work or the work of others) and can be driven to completion by subsequent repeated additions of donor. Our discovery of efficient transfer from simple adenosyl donors removes the limitations imposed by the selectivity of synthetases on the range of molecules that can

be AaT substrates without restricting reaction scale by necessitating the synthesis of oligonucleotide donors. Our yields are comparable to those obtained with pdCpA donors, but pdCpA synthesis requires 6 steps, and overall yields are typically 25 %. We see our methods as complementary to the fully enzymatic synthetase-based methods (Figure 2-1, Top). Fully enzymatic methods do not require prior synthesis and aminoacyl tRNAs are efficient substrates for AaT. On the other hand, our methods can be easily scaled up to modify larger quantities of protein since they do not require the use of purified tRNA.

We have shown that our method can be used to transfer reactive handles to the N-terminus under "protein friendly" (pH ~7, high salt, 37 °C) conditions and that the labeled protein product can be easily purified afterward. Given the previous success of Sisido and co-workers, we expect that further mutation of AaT will allow us to transfer larger side chains that may permit direct transfer of fluorescent probes or affinity tags. We also hope to use the available crystal structures to redesign AaT to act on other N-terminal sequences through a combination of rational design and selection. Finally, our demonstration that AaT can modify α-casein in crude lysates, coupled with the cell-permeability of our moderately polar donors raises the exciting possibility of using AaT as part of an in vivo labeling strategy.

Section 2.3 – Materials and Methods

2.3.1 – Materials

General Information. Chloroacetonitrile, *N,N*-diisopropylethylamine (DIPEA), 5′-*O*-(4,4′-Dimethoxytrityl) adenosine ((DMT)-A), tetrabutylammonium acetate (TBAAc), trifluoroacetic acid (TFA), and triisopropyl silane (TIPSH) were purchased from Sigma-Aldrich (St. Louis, MO). Note: (DMT)-A production was discontinued by Sigma-Aldrich, after which it was purchased as a custom order from ChemGenes Corporation (Wilmington, MA) and was additionally synthesized in-house following the protocol outlined by Ogilvie *et al*.[139] *N*-Boc-L-phenylalanine-*O*-succinimide (Boc-Phe-OSu) and Lysylalanylaminomethylcoumarin (LysAlaAcm) were purchased from Bachem (Torrence, CA). *N*-Boc-L-phenylalanine (Boc-Phe-OH) and all solvents were purchased from Fisher Scientific (Pittsburgh, PA). All deuterated solvents were purchased from Cambridge Isotopes Laboratories, Inc. (Andover, MA). *E. coli* BL21(DE3) cells were purchased from Stratagene (La Jolla, CA). The pEG6 plasmid, containing His_{10}-tagged *E. coli* AaT, was a gift from Alexander Varshavsky (California Institute of Technology). Sequencing-grade trypsin was purchased from Promega (Madison, WI). All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). *N*-Boc-L-leucine (Boc-Leu-OH), *N*-Boc-L-*p*-azidophenylalanine (Boc-Azf-OH), *N*-Boc-L-naphthylalanine (Boc-Nap-OH), *N*-Boc-, *N*-Me-L-phenylalanine (Boc-Mef-OH), *N*-Acetyl-L-phenylalanine (Acf-OH), and *N*-Fmoc-methoxycoumarinylalanine (Fmoc-Mcm-OH) were purchased from Bachem (Torrence, CA). All solvents were purchased from Fisher Scientific (Pittsburgh, PA). QuikChange® site-directed mutagenesis kits were purchased from Stratagene (La Jolla, CA). DNA oligomers were purchased from Integrated DNA

Technologies, Inc (Coralville, IA). Bradford reagent assay kits were purchased from BioRAD (Hercules, CA). Protease inhibitor cocktail was purchased from Sigma-Aldrich (St. Louis, MO). All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). Infrared spectra were recorded on a Nicolet 6700 FT-IR instrument. Milli-Q filtered (18 MΩ) water was used for all aqueous solutions (Millipore; Billerica, MA). Matrix-assisted laser desorption ionization (MALDI) mass spectra were collected using a Bruker Ultraflex III MALDI-TOF-TOF mass spectrometer (Billerica, MA). UV absorbance spectra were obtained with a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA). Donor molecule purification was conducted on a BioCad Sprint FPLC (GMI Inc.; Ramsey, MN; originally from Perseptive Biosystems) with a Waters Sunfire Prep C18-prep OBD column, 5 μm, 17 - 150 mm (Milford, MA). Analytical HPLC assays were performed on an Agilent 1100 HPLC using a Waters Symmetry Shield C18 column. NMR spectra, ^{1}H and ^{13}C, were collected with a Bruker DRX 500 MHz instrument. "Low resolution" electrospray ionization (ESI) mass spectra (LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. "High resolution" ESI mass spectra (HRMS) were obtained on a Waters LCT Premier XE LC/MS. DNA sequencing was performed at the University of Pennsylvania DNA sequencing facility. Fluorescence spectra were collected with a Varian Cary Eclipse fluorescence spectrophotometer fitted with a Peltier multicell holder (currently Agilent Technologies).

2.3.2 – Donor (4a) Synthesis

The synthesis of Phe donor **4a** is given as a general procedure, starting from commercially available Boc-Phe-OH or Boc-Phe-OSu. The carboxy-terminus of the amino acid was activated as a cyanomethyl ester using chloroacetonitrile to give **1a**. Next, the protected amino acid was attached to one of the available hydroxyls (2′ OH or 3′ OH) of (DMT)-A without preference. The resulting protected adenylate (**3a**) was deprotected using a 50/50 mixture of TFA and THF with TIPSH present as a scavenger. The final product, deprotected adenylate (**4a**), was purified on C18 reverse-phase HPLC column.

(S)-Cyanomethyl 2-((tert-butoxycarbonyl)amino)-3-phenylpropanoate (Boc-Phe-OCH_2CN, 1a).

Boc-Phe-OH (2.01 g, 7.57 mmol) was dissolved in 10 mL tetrahydrofuran under ambient conditions. Ten equivalents of chloroacetonitrile (4.78 mL, 75.4 mmol) and 1.1 equiv. of DIPEA (8.30 mmol, 1.45 mL) were added to the reaction mixture and allowed to stir overnight. **1a** was purified on silica gel (20 % EtOAc in hexanes) to give 2.474 g (100 %) of a pale-yellow oil after evaporation. R_f 0.4 in 20 % EtOAc in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.26 (t, J = 7.0 Hz, 2H), 7.20 (t, J = 7.1 Hz, 1H), 7.11 (d, J = 7.3 Hz, 2H), 5.11 (d, J = 7.8 Hz, 1H), 4.69 - 4.57 (m, 3H), 3.09 - 2.98 (m, 2H), 1.36 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.8, 155.1, 135.4, 129.2, 128.7, 127.3, 114.0, 80.9, 80.2, 54.4, 48.8, 37.8, 28.2; HRMS (ESI) calcd m/z for $C_{16}H_{20}N_2O_4Na$ $[M + Na]^+$ 327.132, found 327.133.

(S)-(2R,3R,4R,5R)-2-(6-Amino-9H-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((tert-butoxycarbonyl)amino)-3-phenylpropanoate (Boc-Phe-(DMT)-A, 3a).

(DMT)-A (25 mg, 44 μmol) was dissolved in 5 mL of tetrahydrofuran (dried with 4 Å molecular sieves) in an oven-dried 4 dram reaction vessel. Twenty equivalents of **1a** (272 mg, 0.895 mmol) were added to the reaction mixture. TBAAc (4.9 mg, 16 μmol), the reaction catalyst, was added last to the reaction mixture. In reproductions of this reaction, TBAAc additions varied according to reaction progress. The reaction stirred under argon overnight at room temperature for 1 - 4 days. After evaporation of solvent, the product was purified on silica gel (gradient solvent system: 20 – 0 % petroleum ether in EtOAc, 5 % methanol in EtOAc) to give 27 mg (75 %) of a pale-yellow foam after evaporation. R_f 0.5 in 5 % methanol in EtOAc. ^{1}H NMR (500 MHz, THF-d_8) δ 8.08 (s, 1H), 8.00 (s, 1H), 7.44 (d, J = 3.68Hz, 2H), 7.32 (d, J = 8.85 Hz, 4H), 7.20 (dd, J = 8.33Hz, J = 6.66, 2H), 7.13 (t, J = 7.25Hz, 1H), 7.21 - 7.10 (m, 4H), 6.77 (d, J = 4.30 Hz, 2H), 6.75 (d, J = 4.29 Hz, 2H), 5.92 (d, J = 4.87 Hz, 1H), 5.49 (t, J = 4.34 Hz, 1H), 5.13 (t, J = 5.54 Hz, 1H), 4.51 - 4.47 (m, 1H), 4.14 (t, J = 4.07Hz, 1H), 3.71 (s, 6H) (S), 3.42 - 3.32 (m, 2H), 3.12 - 2.96 (m, 3H), 1.37 (s, 9H); ^{13}C NMR (125 MHz, THF-d_8) δ 172.2, 159.9, 157.6, 156.7, 153.8, 146.3, 138.3, 146.3, 138.3, 1367.0, 136.9, 131.2, 131.1, 130.3, 129.3, 129.2, 129.1, 128.6, 128.5, 127.5, 114.0, 113.9, 89.7, 87.5, 82.5, 79.7, 75.1, 73.5, 64.7, 56.4, 55.5, 38.9, 30.6, 28.8; LRMS (ESI) calcd m/z for $C_{45}H_{48}N_6O_9H$ $(M + H)^+$ 816.9, $C_{45}H_{48}N_6O_9Na$ $(M + Na)^+$ 839.3, found 817.3, 839.3.

(S)-(2R,3R,4R,5R)-2-(6-Amino-9H-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((tert-butoxycarbonyl)amino)-3-phenylpropanoate (3a) from 2a.

(DMT)-A (26 mg, 45 μmol) was dissolved in 5 mL of tetrahydrofuran (dried with 4 Å molecular sieves) in an oven-dried 4-dram reaction vessel. Twenty equivalents of **2a** (321 mg, 0.885 mmol) were added to the reaction mixture. TBAAc (19 mg, 64 μmol), the reaction catalyst, was added last to the reaction mixture in two separate additions, 6 mg, at the beginning of the reaction, and 13 mg after the first 24 h of reaction. The reaction stirred under argon overnight at room temperature for 2 days. After evaporation of solvent, the product was purified on silica gel (gradient solvent system: 20 – 0 % petroleum ether in EtOAc, 5 % methanol in EtOAc) to give 19.1 mg (52.1 %) of a pale-yellow foam after evaporation. LRMS (ESI) calcd m/z for $C_{45}H_{48}N_6O_9H(M+H)^+$ 816.9, $C_{45}H_{48}N_6O_9Na$ $(M + Na)^+$ 839.3, found 817.4, 839.4.

(S)-(2R,3R,4R,5R)-2-(6-Amino-9H-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-phenylpropanoate (Phe-A, 4a).

3a (50 mg) was dissolved in 1 mL of THF and 1 mL trifluoroacetic acid (TFA) and reacted overnight with four equiv. of TIPSH. Upon addition of TFA, the reaction turned a bright-orange color and at the completion of the reaction, the color had become a dark brown. The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a 1 mL water final back-extraction against the dichloromethane layer. The water-soluble layer containing **4a**

was then HPLC-purified on a C18-prep column using an increasing acetonitrile gradient (Gradient 2: 1 – 100 % acetonitrile with water and 0.1 % TFA, 1 % per min) to remove any protected material. Prep. HPLC purified to give 13.6 mg (53.5 %) of a white solid after evaporation. MALDI MS calcd m/z for $C_{19}H_{22}N_6O_5H$ $(M + H)^+$ 415.4, $C_{19}H_{22}N_6O_5Na$ $(M + Na)^+$ 437.4, found 415.0, 437.0.

2.3.3 – Donor (4b-i) Synthesis

As described above, the carboxy-terminus of the amino acid (or analog) was activated as a cyanomethyl ester using chloroacetonitrile to give **1b-i**. Next, the protected amino acid was attached to one of the available hydroxyls (2′ OH or 3′ OH) of DMT-A without preference. The resulting protected adenylate (**3b-i**) was deprotected using a 50/50 solvent mixture of trifluoroacetic acid and THF (1 mL) or neat TFA (1 mL) with TIPSH present as a scavenger, stirred for 24 h, concentrated under reduced pressure, extracted using DCM and water as mentioned in the main text, and HPLC purified. The final product, deprotected adenylate (**4b-i**), was purified *via* C18 column purification using an HPLC.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-4-methylpentanoate (Boc-Leu-OCH_2CN, 1b).

Chloroacetonitrile (5 mL) and DIPEA (160 mg, 0.21 mL, 1.2 mmol) were added to Boc-Leu-OH (252 mg, 1.09 mmol) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (20 % ethyl acetate in hexanes) afforded 284 mg of a pale yellow oil in 96 % yield. R_f 0.5 in 20 % ethyl acetate in

hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ 5.00 (d, J = 7.7 Hz, 1H), 4.80 – 4.65 (m, 2H), 4.30 – 4.29 (m, 1H), 1.70 – 1.65 (m, 1H), 1.60 – 1.47 (m, 2H), 1.39 (s, 9H), 0.91 – 0.89 (m, 6H); ^{13}C NMR (125 MHz, $CDCl_3$): δ172.2, 155.5, 114.21, 80.3, 53.2, 52.0, 41.0, 28.3, 24.8, 22.8, 21.7; HRMS (ESI) m/z calcd for $C_{13}H_{22}N_2O_4Na$ $[M + Na]^+$ 293.147, found 293.151.

(*S*)-cyanomethyl 3-(4-azidophenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (Boc-Azf-OCH$_2$CN, 1c).

Chloroacetonitrile (5 mL) and DIPEA (99 mg, 0.13 mL, 0.76 mmol) were added to Boc-Azf-OH (212 mg, 0.691 mmol) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (30 % ethyl acetate in hexanes) afforded 230 mg of a pale yellow oil in 96 % yield. R_f 0.5 in 30 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.13 (d, J = 8.3 Hz, 2H), 6.97 (d, J = 8.4 Hz, 2H), 5.00 (d, J = 7.7 Hz, 1H), 4.80 – 4.65 (m, 2H), 4.62 – 4.58 (m, 1H), 3.12 – 3.01 (m, 2H), 1.40 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 170.7, 155.1, 139.4, 132.0, 130.7, 119.5, 114.0, 80.7, 54.4, 49.0, 37.5, 28.3; HRMS (ESI) m/z calcd for $C_{16}H_{19}N_5O_4Na$ $[M + Na]^+$ 368.133, found 368.135.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-3-(naphthalen-2-yl)propanoate (Boc-Nap-OCH$_2$CN, 1d).

Chloroacetonitrile (2.5 mL) and DIPEA (25 mg, 33 μL, 0.19 mmol) were added to Boc-NapAla-OH (54.8 mg, 0.174 mmol) and stirred for 12 h. The solvent was

removed under reduced pressure and SiO_2 flash chromatography (30 % ethyl acetate in hexanes) afforded 61 mg of a pale yellow solid in 99 % yield. R_f 0.5 in 30 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.83-7.81 (m, 3H), 7.62 (s, 1H), 7.51 – 7.46 (m, 2H), 7.29 (dd, J = 8.4, 1.5 Hz, 1H), 5.00 (d, J = 7.4 Hz, 1H), 4.79 – 4.64 (m, 3H), 3.31 (dd, J = 14.1, 5.9 Hz), 3.25 (dd, J = 14.0, 6.5 Hz, 1H), 1.41 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 170.9, 155.2, 133.6, 132.8, 132.8, 128.8, 128.3, 127.8, 127.2, 126.5, 126.2, 114.0, 80.7, 54.5, 49.0, 38.2, 28.4; HRMS (ESI) m/z calcd for $C_{20}H_{22}N_2O_4Na$ $[M + Na]^+$ 377.147, found 377.149.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)(methyl)amino)-3-phenylpropanoate (Boc-Mef-OCH$_2$CN, 1e).

Chloroacetonitrile (1.1 mL) and DIPEA (258 mg, 0.341 mL, 1.99 mmol) were added to Boc-MePhe-OH (497 mg, 1.78 mmol) in tetrahydrofuran (5 mL) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (20 – 35 % ethyl acetate in hexanes) afforded 505 mg of a pale yellow oil in 89 % yield. R_f 0.3 in 20 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.28 – 7.25 (m, 2H), 7.20 – 7.16 (m, 3H), 4.72 – 4.68 (m, 2H), 4.53 – 4.52 (m, 1H), 3.27 (m, 1H), 3.11 – 3.02 (m, 1H), 2.67 (d, J = 17 Hz, 3H), 1.36 (d, J = 12 Hz, 9H)**;** ^{13}C NMR (125 MHz, $CDCl_3$): δ 169.9, 155.5, 136.8, 128.9, 128.6, 128.4, 126.8, 126.7, 114.2, 80.8, 61.5, 48.7, 35.4, 33.0, 28.1; HRMS (ESI) m/z calcd for $C_{17}H_{22}N_2O_4Na$ $[M + Na]^+$ 341.147, found 341.149.

(*S*)-cyanomethyl 2-acetamido-3-phenylpropanoate (Acf-OCH$_2$CN, 1f).

Chloroacetonitrile (2 mL) and DIPEA (630 mg, 0.841 mL, 4.87 mmol) were added to *N*-Ac-Phe-OH (501 mg, 2.42 mmol) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (50 % ethyl acetate in hexanes) afforded (535 mg) of a pale yellow oil in 90 % yield. R_f 0.2 in 50 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.26 (t, J = 7 Hz, 2H), 7.22 (t, J = 2.3 Hz, 1H), 7.07 (d, J = 7.1 Hz, 2H), 6.02 (d, J = 7.4 Hz, 1H), 4.83 (dd, J = 14.0, 6.5 Hz, 1H), 4.66 (dd, J = 15.7, 4.7 Hz, 2H), 3.10 – 3.01 (m, 2H), 1.91 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 170.6, 170.1, 135.2, 129.3, 139.0, 127.7, 114.0, 53.6, 53.2, 49.0, 37.7, 23.0; HRMS (ESI) m/z calcd for $C_{13}H_{14}N_2O_3Na$ $[M + Na]^+$ 269.090, found 269.091.

(*S*)-cyanomethyl 2-azido-3-phenylpropanoate (Boc-N$_3$f-OCH$_2$CN, 1g).

Chloroacetonitrile (1 mL) and DIPEA (310 mg, 0.411 mL, 2.41 mmol) were added to 2-azido-3-phenylpropanoate and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (40 – 50 % ethyl acetate in hexanes) afforded (250 mg) of a white solid in 50 % yield. R_f 0.5 in 50 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (t, J = 7.6 Hz, 2H), 7.32-7.29 (m, 1H), 7.25 (t, J = 7.6 Hz, 2H), 4.77 (t, J = 1.0 Hz, 1H), 4.18 (dd, J = 6.2, 1.8 Hz, 1H), 3.21 (dd, J = 14.0, 5.8 Hz, 1H), 3.08 (dd, J = 13.9, 8.4 Hz, 1H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 168.8, 135.1, 129.3, 129.1, 127.8, 113.7, 62.9, 49.3, 37.7; HRMS (ESI) m/z calcd for $C_{11}H_{10}N_4O_2Na$ $[M + Na]^+$ 253.070, found 253.071.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-3-(7-methoxy-2-oxo-2*H*-chromen-4-yl) propanoate (Boc-Mcm-OCH_2CN, 1h).

A solution of 20 % piperidine in tetrahydrofuran (4 mL) was added to Fmoc-Mcm-OH (200 mg, 0.439 mmol) and was stirred for 15 minutes. The solvent was removed under reduced pressure and ethyl acetate (10 mL) and 2M sodium hydroxide was added until pH > 8 followed by addition of Boc anhydride (110 mg, 0.526 mmol) to the white residue and stirred overnight. The organic layer was extracted and washed with saturated sodium bicarbonate (2 x 10 mL) and all aqueous fractions were combined and acidified with 1 M sodium bisulfate until pH < 2. The aqueous layer was then extracted with ethyl acetate (3 x 10 mL), dried with magnesium sulfate, and concentrated under reduced pressure. Chloroacetonitrile (1 mL) and DIPEA (62 mg, 82 μL, 0.48 mmol) were added to the white residue and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (50 % - 65 % ethyl acetate in hexanes) afforded 56 mg of a pale yellow oil in 35 % yield. R_f 0.4 in 50 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (d, J = 8.7 Hz, 1H), 6.90 (dd, J = 8.9, 2.4 Hz, 1H), 6.85 (d, J = 2.1 Hz, 1H), 6.12 (s, 1H), 5.16 (d, J = 7.7 Hz, 1H), 4.79 (d, J = 4.2 Hz, 2H), 4.69-4.68 (m, 1H), 3.88 (s, 3H), 3.32 (dd, J = 14.1, 5.2 Hz, 1H), 3.12 (dd, J = 13.9, 8.4 Hz, 1H), 1.41 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.0, 163.2, 160.8, 155.9, 155.1, 150.5, 125.3, 113.7, 113.3, 113.0, 112.3, 101.6, 81.3, 56.0, 52.8, 49.5, 34.6, 28.4; HRMS (ESI) m/z calcd for $C_{20}H_{22}N_2O_7Na$ $[M + Na]^+$ 425.133, found 425.133.

(*S*)-cyanomethyl 3-(4-benzoylphenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (Boc-Bzf-OCH_2CN, 1h).

Chloroacetonitrile (1 mL) and DIPEA (20 mg, 34 μL, 0.21 mmol) were added to Boc-*p*-benzoylphenylalanine (61 mg, 0.19 mmol). The reaction was stirred for 18 h, then concentrated under reduced pressure. SiO_2 flash chromatography (40 % ethyl acetate in hexanes) afforded 70 mg of a pale yellow oil in 93 % yield. R_f 0.5 in 50 % ethyl acetate in hexanes. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.79 (dd, *J* = 7.1, 4.9 Hz, 4H), 7.59 (t, *J* = 7.6 Hz, 1H), 7.48 (t, *J* = 7.6 Hz, 2H), 7.29 (d, *J* = 8.0 Hz, 2H), 5.00 (d, *J* = 7.4 Hz, 1H), 4.83 (d, *J* = 15.7 Hz, 1H), 4.72 – 4.69 (m, 2H), 3.25 (dd, *J* = 13.7, 5.8 Hz, 1H), 3.16 (dd, *J* = 13.1, 6.5 Hz, 1H), 1.43 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 196.4, 170.6, 155.1, 140.3, 137.6, 136.9, 132.7, 130.8, 130.2, 129.4, 128.5, 113.9; HRMS (ESI) m/z calcd for $C_{23}H_{24}N_2NaO_5$ [M + Na]+ 431.158, found 431.158.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl) methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-4-methylpentanoate (Boc-Leu-(DMT)-A, 3b).

Tetrahydrofuran (5 mL) was added to Boc-Leu-OCH_2CN **1b** (226 mg, 0.837 mmol), (DMT)-A (120 mg, 0.211 mmol), and TBAAc (15 mg, catalyst) and stirred for 24 h. The solvent was removed under reduced pressure and chromatography (80 – 100 % ethyl acetate in petroleum ether, 5 % methanol in ethyl acetate), afforded 122 mg of a white solid in 74 % yield. R_f 0.5 in 5 % methanol in ethyl acetate; ^{1}H and ^{13}C NMR for **3b** shown below; LRMS (ESI) m/z calcd for $C_{42}H_{50}N_6O_9Na$ $(M + Na)^+$ 805.4, found

805.3.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 3-(4-azidophenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (Boc-Azf-(DMT)-A, 3c).

Tetrahydrofuran (5 mL) was added to Boc-Azf-OCH_2CN **1c** (101 mg, 0.290 mmol), (DMT)-A (43 mg, 75 μmol), and TBAAc (2 mg, catalyst) and stirred for 24 h. Solvent was removed under vacuum and preparative TLC (100 % ethyl acetate), afforded 44 mg of white solid in 69 % yield. R_f 0.3 - 0.5 in ethyl acetate; ^1H and ^{13}C NMR for **3c** shown below; LRMS (ESI) m/z calcd for $C_{42}H_{47}N_9O_9Na$ $(M + Na)^+$ 880.3, found 880.4.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl) methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(naphthalen-2-yl)propanoate (Boc-Nap-(DMT)-A, 3d).

Tetrahydrofuran (5 mL) was added to Boc-Nap-OCH_2CN **1d** (58 mg, 0.16 mmol), (DMT)-A (26 mg, 46 μmol), and TBAAc (6.4 mg, catalyst) and stirred for 24 h. The solvent was removed under reduced pressure and preparative TLC (5 % methanol in ethyl acetate), afforded 36 mg of a white solid in 91 % yield. R_f 0.3 - 0.5 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for **3d** shown below; LRMS (ESI) m/z calcd for $C_{49}H_{50}N_6O_9Na$ $(M + Na)^+$ 889.9, found 889.4.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl) methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)(methyl)amino)-3-phenylpropanoate (Boc-Mef-(DMT)-A, 3e).

Tetrahydrofuran (3.5 mL) and DIPEA (44 mg, 60 μL, 0.35 mmol) were added to Boc-Mef-OCH_2CN **1e** (110 mg, 0.347 mmol), (DMT)-A (50 mg, 89 μmol), and TBAAc (3 mg, catalyst) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (50 – 100 % ethyl acetate in hexanes), afforded 10 mg of a white solid in 14 % yield. R_f 0.4 - 0.5 in ethyl acetate; ^1H and ^{13}C NMR for **3e** shown below; LRMS (ESI) m/z calcd for $C_{46}H_{50}N_6NaO_9$ $(M + Na)^+$ 853.4, found 853.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl) methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-acetamido-3-phenylpropanoate (Acf-(DMT)-A, 3f).

Tetrahydrofuran (5 mL) and DIPEA (250 mg, 0.30 mL, 1.9 mmol) were added to Boc-Acf-OCH_2CN **1f** (400 mg, 1.93 mmol), (DMT)-A (73 mg, 0.13 mmol) and TBAAc (5 mg, catalyst) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (5 % methanol in ethyl acetate), afforded 75 mg of a clear solid that was mix of DMTA and product. Subsequent HPLC analysis suggests approximately 38 % of this mixture was product. R_f 0.1 - 0.25 % methanol in ethyl acetate; ^1H and ^{13}C NMR for **3f** shown below; LRMS (ESI) m/z calcd for $C_{42}H_{43}N_6O_8$ $(M + H)^+$ 759.3 $C_{42}H_{42}N_6NaO_8$ $(M + Na)^+$ 781.3, found 759.3, 781.4.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-azido-3-phenylpropanoate (Boc-N_3f-(DMT)-A, 3g).

Tetrahydrofuran (3.5 mL) and DIPEA (44 mg, 60 µL, 0.35 mmol) were added to Boc-N_3f-OCH_2CN **1g** (80 mg, 0.35 mmol), (DMT)-A (50 mg, 89 µmol), and TBAC (3 mg, catalyst) and stirred for 12 h. The solvent was removed under reduced pressure and purification by silica flash chromatography (75 – 100 % ethyl acetate in hexanes), afforded 53 mg of a white solid in 80 % yield. R_f 0.3 - 0.5 in ethyl acetate; ^1H and ^{13}C NMR for **3g** shown below; LRMS (ESI) m/z calcd for $C_{40}H_{39}N_8O_7$ $(M + H)^+$ 743.3, $C_{40}H_{39}N_8O_7Na$ $(M + Na)^+$ 765.3, found 743.4, 765.4; FTIR (Film) *ν*max 3334, 3179, 3063, 3032, 2953, 2934, 2837, 2112 ($-N_3$), 1757, 1644, 1606, 1508, 1252, 1177 cm^{-1}.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl) methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(7-methoxy-2-oxo-2*H*-chromen-4-yl)propanoate (Boc-Mcm-(DMT)-A, 3h).

Tetrahydrofuran (3.5 mL) and DIPEA (44 mg, 60 µl, 0.35 mmol) were added to Boc-Mcm-OCH_2CN **1h** (56 mg, 0.15 mmol), (DMT)-A (50 mg, 89 µmol), and TBAAc (3 mg, catalyst) and stirred for 12 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (50 – 100 % ethyl acetate in hexanes), afforded 35 mg of a white solid in 43 % yield. R_f 0.3 – 0.4 in ethyl acetate; ^1H and ^{13}C NMR for **3h** shown below; LRMS (ESI) m/z calcd for $C_{49}H_{50}N_6NaO_{12}$ $(M + Na)^+$ 937.3, found 937.4.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 3-(4-benzoylphenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (Boc-Bzf-(DMT)-A, 3i).

THF (3.5 mL) and DIPEA (44 mg, 60 μL, 0.35 mmol) were added to Boc-Bzf-OCH_2CN (70 mg, 0.17 mmol), (DMT)-A (50 mg, 89 μmol) and TBAAc (5 mg, catalyst) and stirred for 16 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (75 % – 100 % ethyl acetate in hexanes) afforded 66 mg of a white solid in 81 % yield. R_f 0.3 in 100 % ethyl acetate. LRMS (ESI) m/z calcd for $C_{52}H_{52}N_6NaO_{10}$ $(M+ Na)^+$ 943.4, found 943.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-4-methylpentanoate (Leu-A, 4b).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (99 mg, 0.13 mL, 0.63 mmol) were added to **3b** (123 mg, 0.157 mmol) following the general deprotection procedure. HPLC/MALDI analysis m/z calcd $C_{16}H_{25}N_6O_5$ $(M + H)^+$ 381.2; Gradient 2; retention time 14.9 min, found 381.1; retention time 16.6 min, found 381.1.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-3-(4-azidophenyl)propanoate (Azf-A, 4c).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (33 mg, 42 μL, 0.21 mmol) were added to **3c** (44 mg, 52 μmol) following the general deprotection procedure. HPLC/MALDI analysis m/z calcd $C_{19}H_{22}N_9O_5$ $(M + H)^+$ 456.2; Gradient 2;

retention time 19.6 min, found 456.1; retention time 21.1 min, found 456.1.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-3-(naphthalen-2-yl)propanoate (Nap-A, 4d).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (26 mg, 0.34 mL, 0.17 mmol) were added to **3d** (36 mg, 41 μmol) following the general deprotection procedure. HPLC/MALDI analysis m/z calcd $C_{23}H_{25}N_6O_5$ $(M + H)^+$ 465.2, $C_{23}H_{24}N_6O_5Na$ $(M + Na)^+$ 487.2; Gradient 2; retention time 21.9 min, found 464.6, 486.6; retention time 23.4 min, found 464.6, 486.6.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-(methylamino)-3-phenylpropanoate (Mef-A, 4e).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and oxalic acid (2.9 mg) were added to **3e** (31 mg, 37 μmol) following the general deprotection procedure. No TIPSH was used in this reaction. HPLC/MALDI analysis m/z calcd $C_{20}H_{25}N_6O_5$ $(M + H)^+$ 429.2, $C_{20}H_{24}N_6O_5Na$ $(M + Na)^+$ 451.2; Gradient 2; retention time 16.3 min, found 429.1; retention time 18.6 min, found 429.1, 451.1.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-acetamido-3-phenylpropanoate (Acf-A, 4f).

During HPLC analysis, **3f** was directly deprotected to **4f** by exposure to the 0.1 % TFA in the HPLC solvent and confirmed *via* MALDI. HPLC/MALDI analysis m/z calcd

$C_{21}H_{25}N_6O_6$ $(M + H)^+$ 457.2, $(M + Na)^+$ $C_{21}H_{24}N_6NaO_6$ 479.2; Gradient 2; retention time 18.2 min, found 457.3, 479.3; retention time 19.5 min, found 457.3, 479.3.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-azido-3-phenylpropanoate (N_3f-A, 4g).

Trifluoroacetic acid (1 mL) and TIPSH (21 mg, 30 µl, 0.13 mmol) were added to **3g** (25 mg, 33 µmol) following the general deprotection procedure. HPLC/MALDI analysis m/z calcd $C_{19}H_{21}N_8O_5$ $(M + H)^+$ 441.2, $C_{19}H_{20}N_8NaO_8$ $(M + Na)^+$ 463.1; Gradient 2; retention time 24.5 min, found 440.9, 462.9; retention time 24.7 min, found 440.9, 462.9.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-3-(7-methoxy-2-oxo-2*H*-chromen-4-yl)propanoate (Mcm-A, 4h).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and oxalic acid (3.9 mg) were added to **3e** (31 mg, 37 µmol) following the general deprotection procedure, except was worked up in water only and water-soluble portion was HPLC purified. No TIPSH was used in this reaction. HPLC/MALDI analysis m/z calcd $C_{23}H_{25}N_6O_8$ $(M + H)^+$ 513.2, $C_{23}H_{24}N_6O_8Na$ $(M + Na)^+$ 535.2; Gradient 2; retention time 19.9 min, found 513.1, 535.1; retention time 20.9 min, found 513.1, 535.1.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-3-(4-benzoylphenyl)propanoate (Bzf-A, 4i).

TFA (1 mL) and TIPSH (40 mg, 50 μL, 0.26 mmol) were added to DMT-A benzoylphenylalanine (59 mg, 64 μmol) according the general deprotection procedure. HPLC/MALDI analysis m/z calcd $C_{26}H_{27}N_6O_6$ $(M + H)^+$ 419.2, $C_{26}H_{26}N_6NaO_6$ $(M + Na)^+$ 441.2; Gradient 2, retention time 19.7 min found 519.1, 541.1.

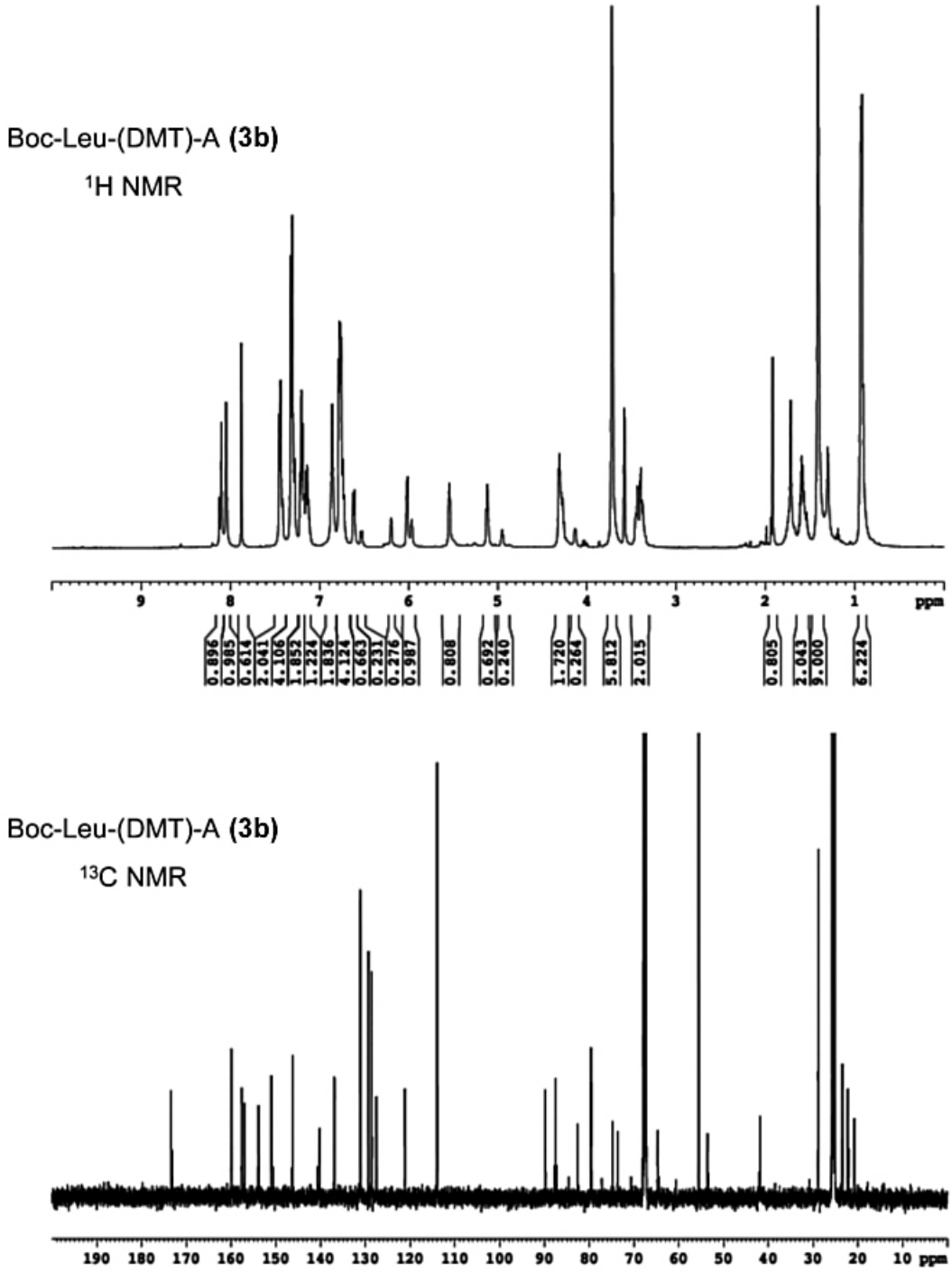


Figure 2-5. ^{1}H and ^{13}C NMR Characterization of Boc-Leu-(DMT)-A (3b)

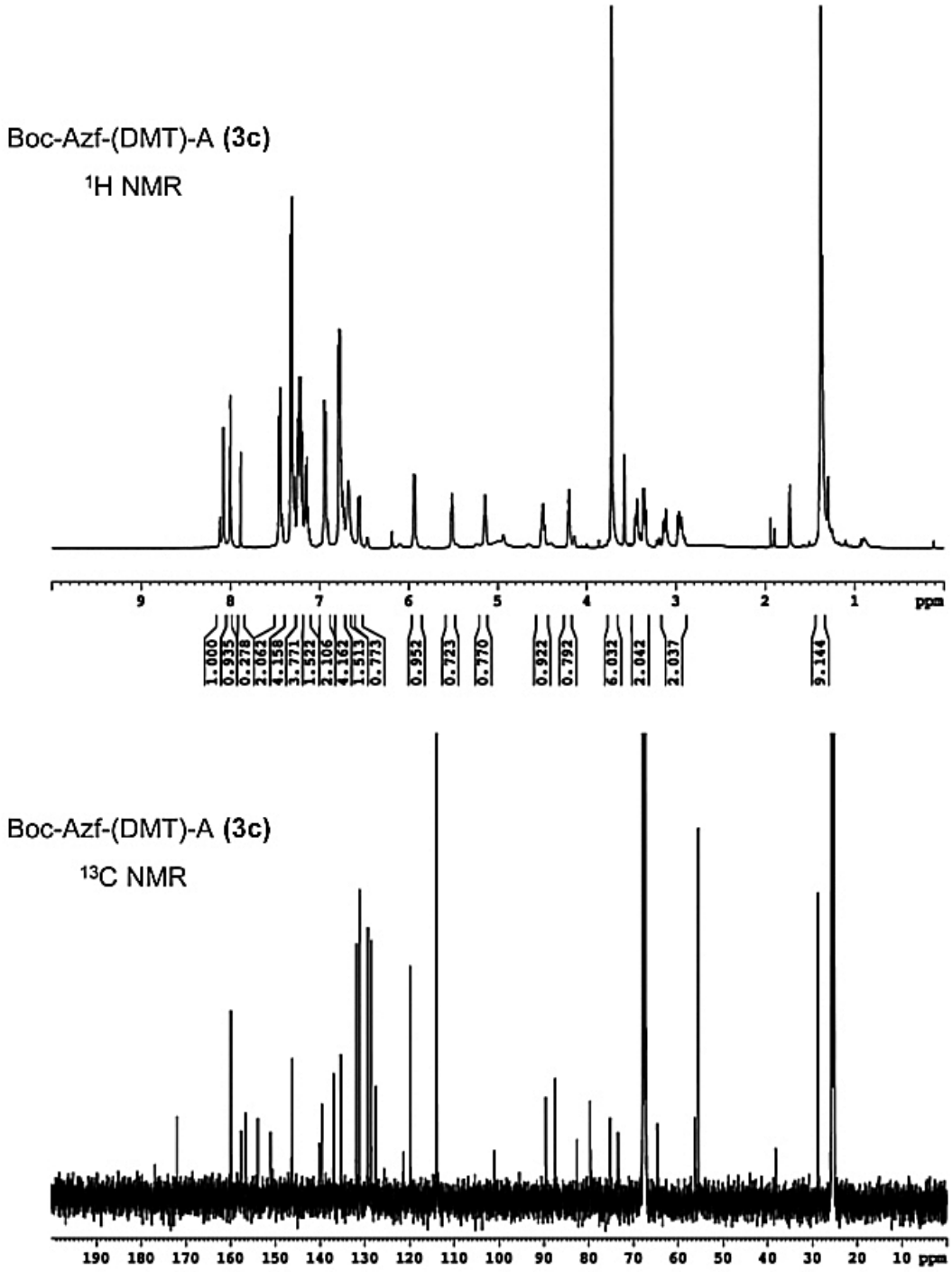


Figure 2-6. ^{1}H and ^{13}C NMR Characterization of Boc-Azf-(DMT)-A (3c)

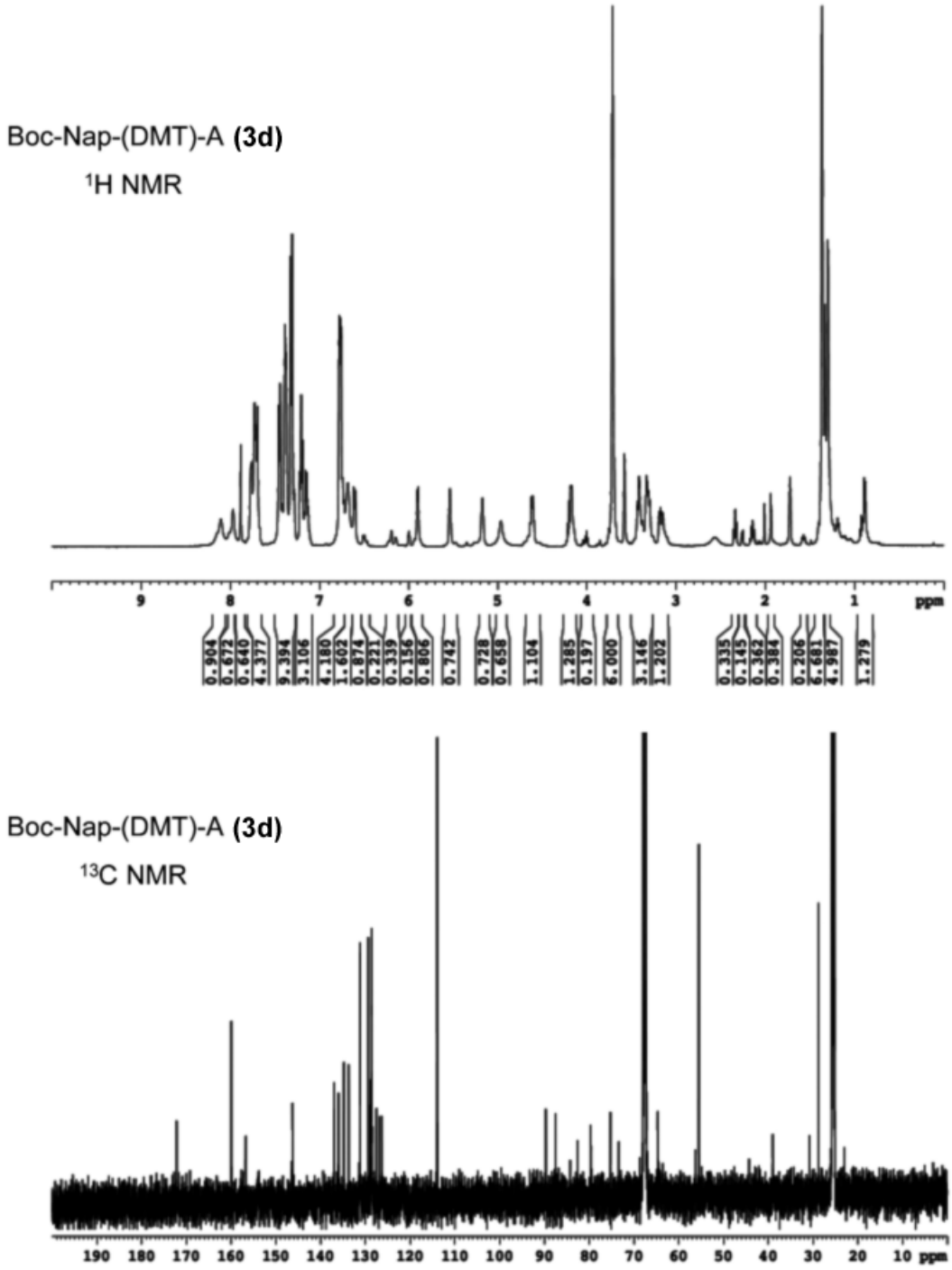


Figure 2-7. ^{1}H and ^{13}C NMR Characterization of Boc-Nap-(DMT)-A (3d)

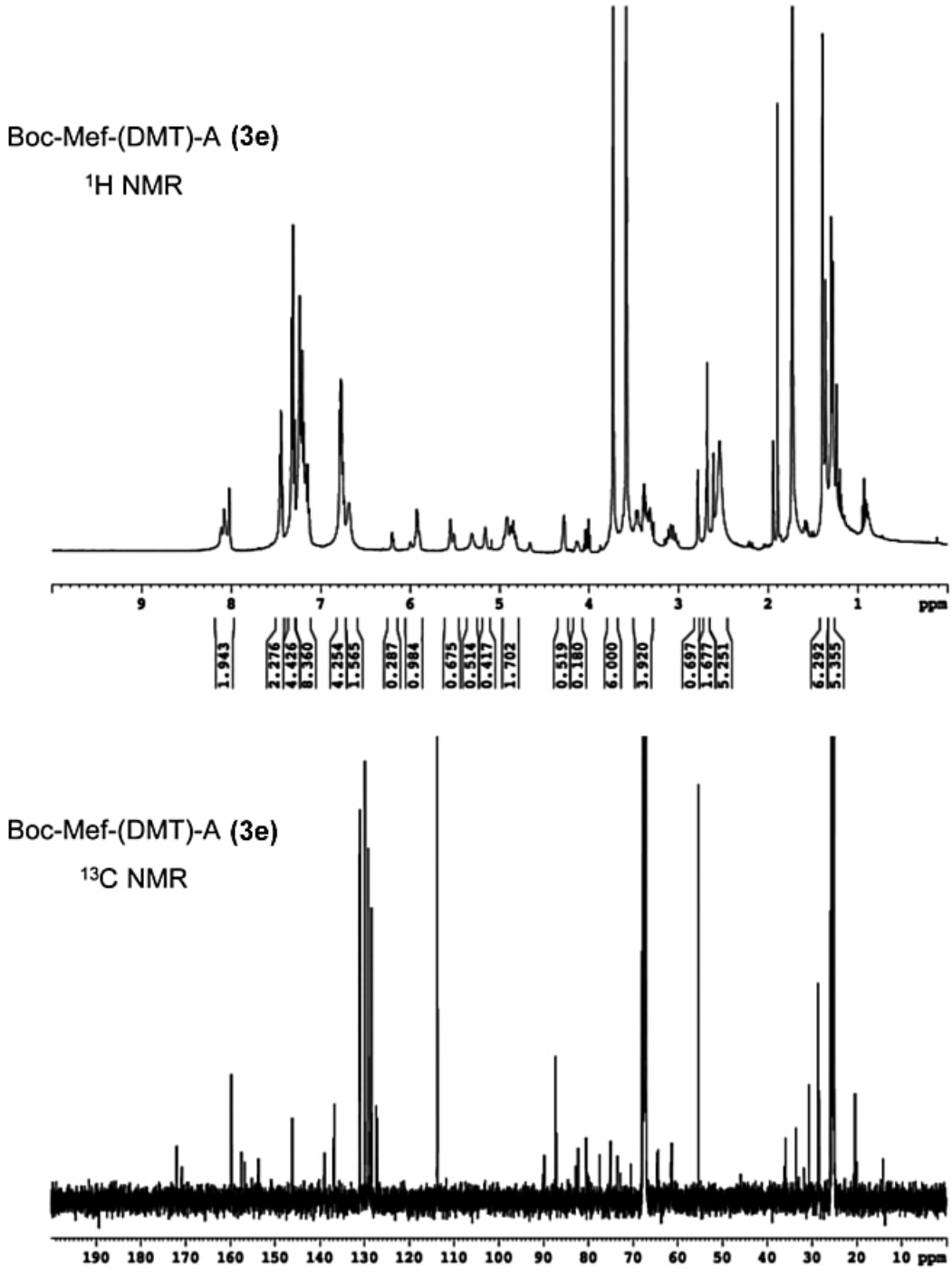


Figure 2-8. ^{1}H and ^{13}C NMR Characterization of Boc-Mef-(DMT)-A (3e)

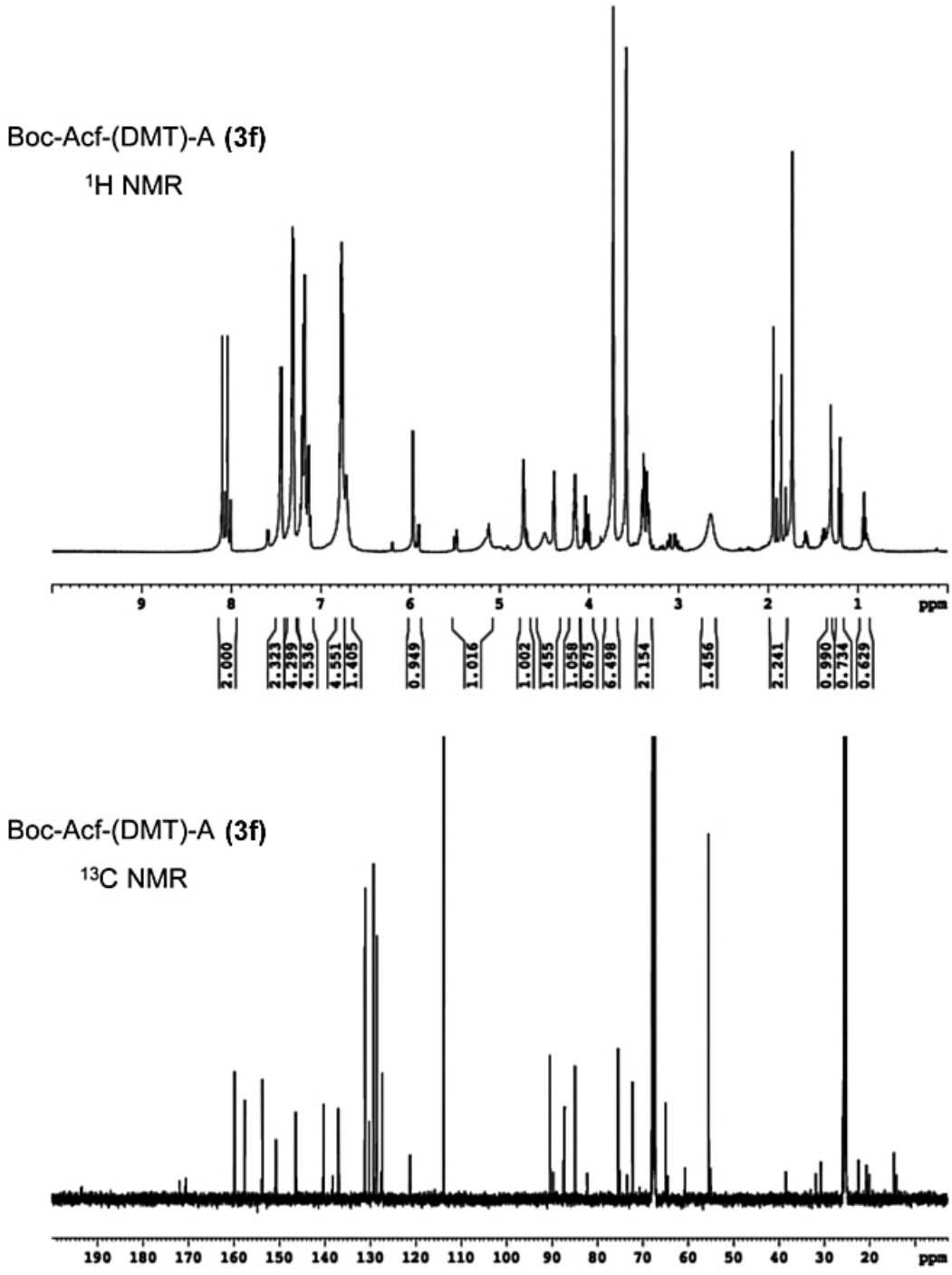


Figure 2-9. ^{1}H and ^{13}C NMR Characterization of Boc-Acf-(DMT)-A (3f)

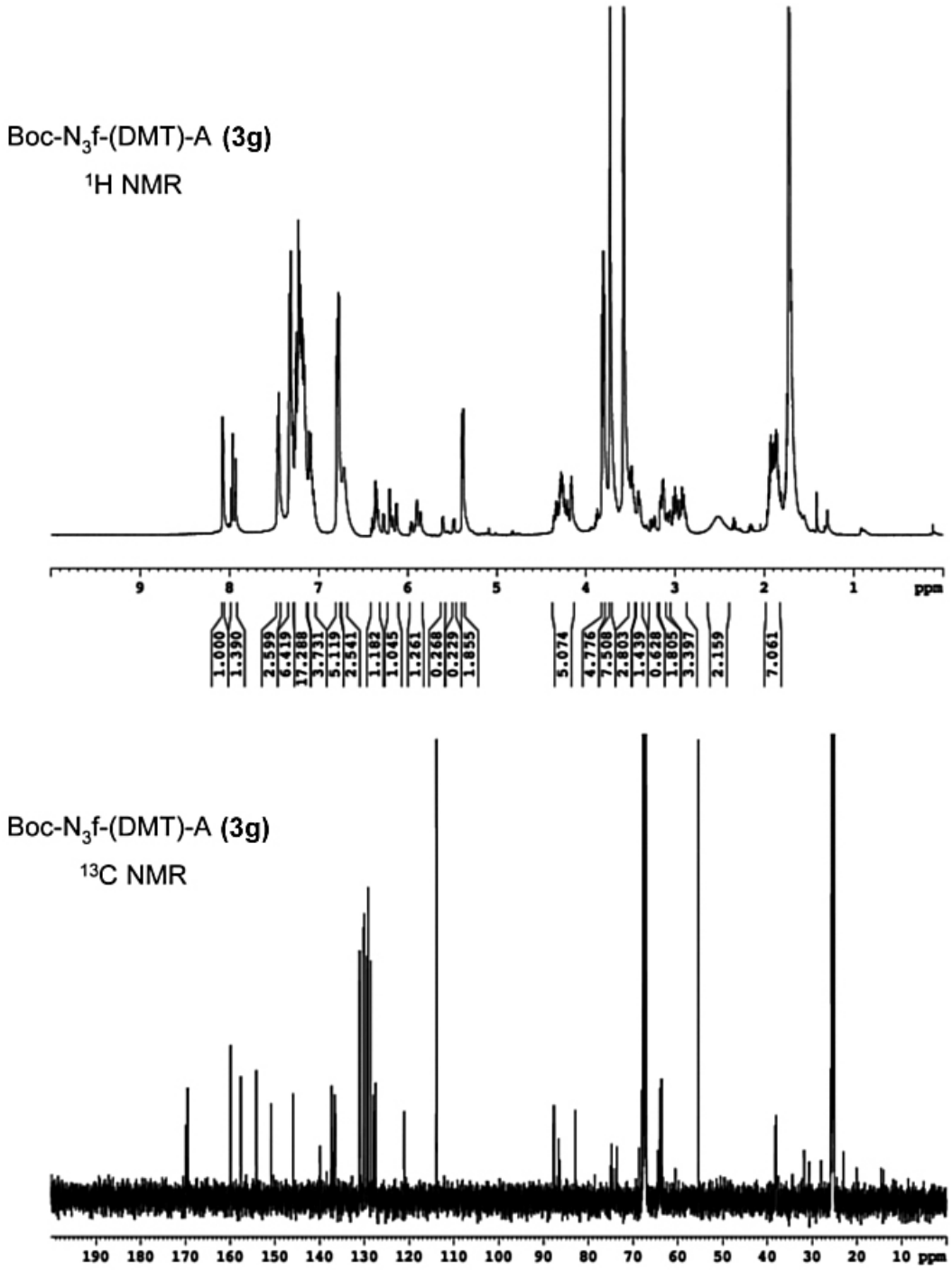


Figure 2-10. ^{1}H and ^{13}C NMR Characterization of Boc-N_3f-(DMT)-A (3g)

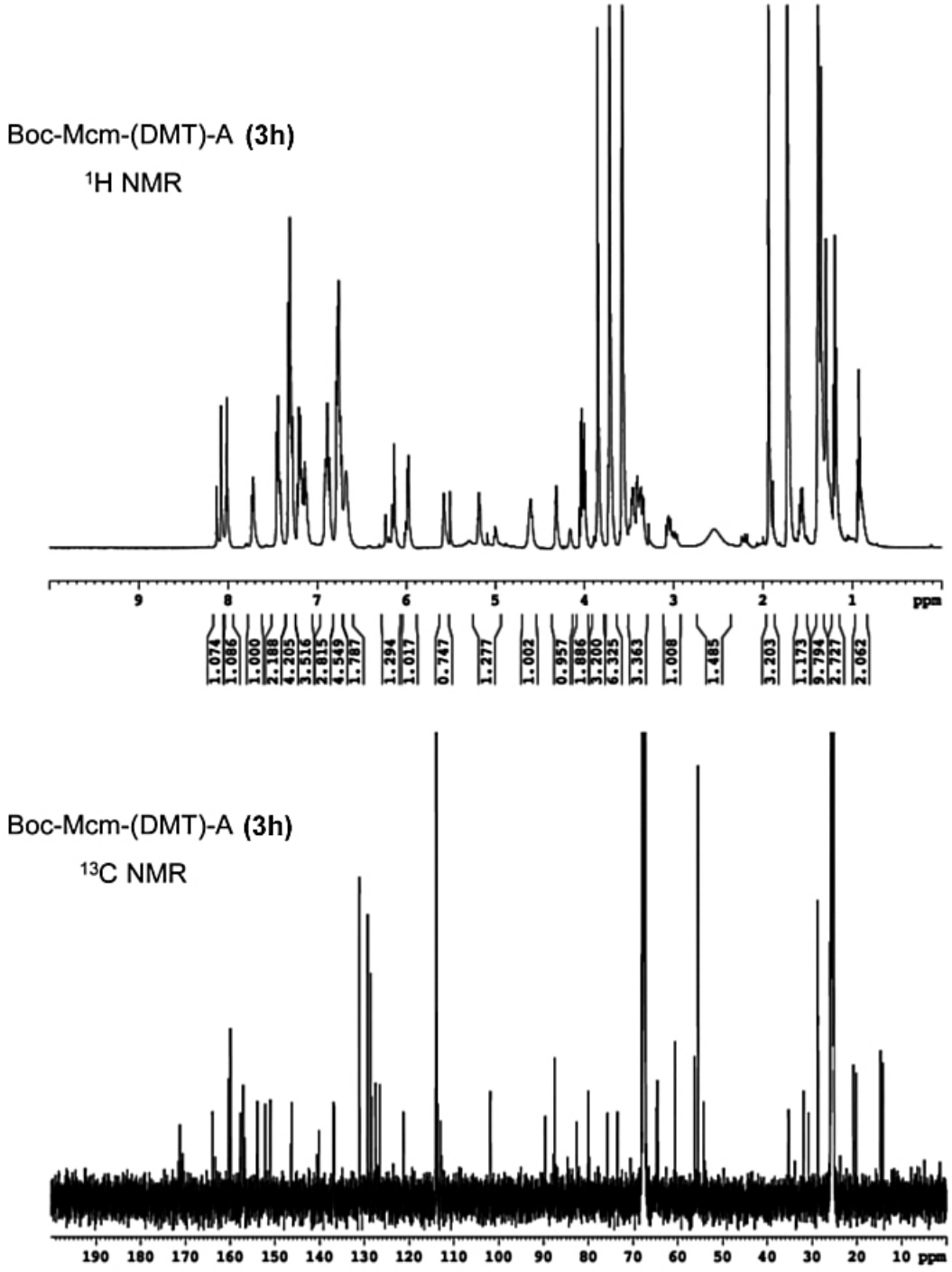


Figure 2-11. ^{1}H and ^{13}C NMR Characterization of Boc-Mcm-(DMT)-A (3h)

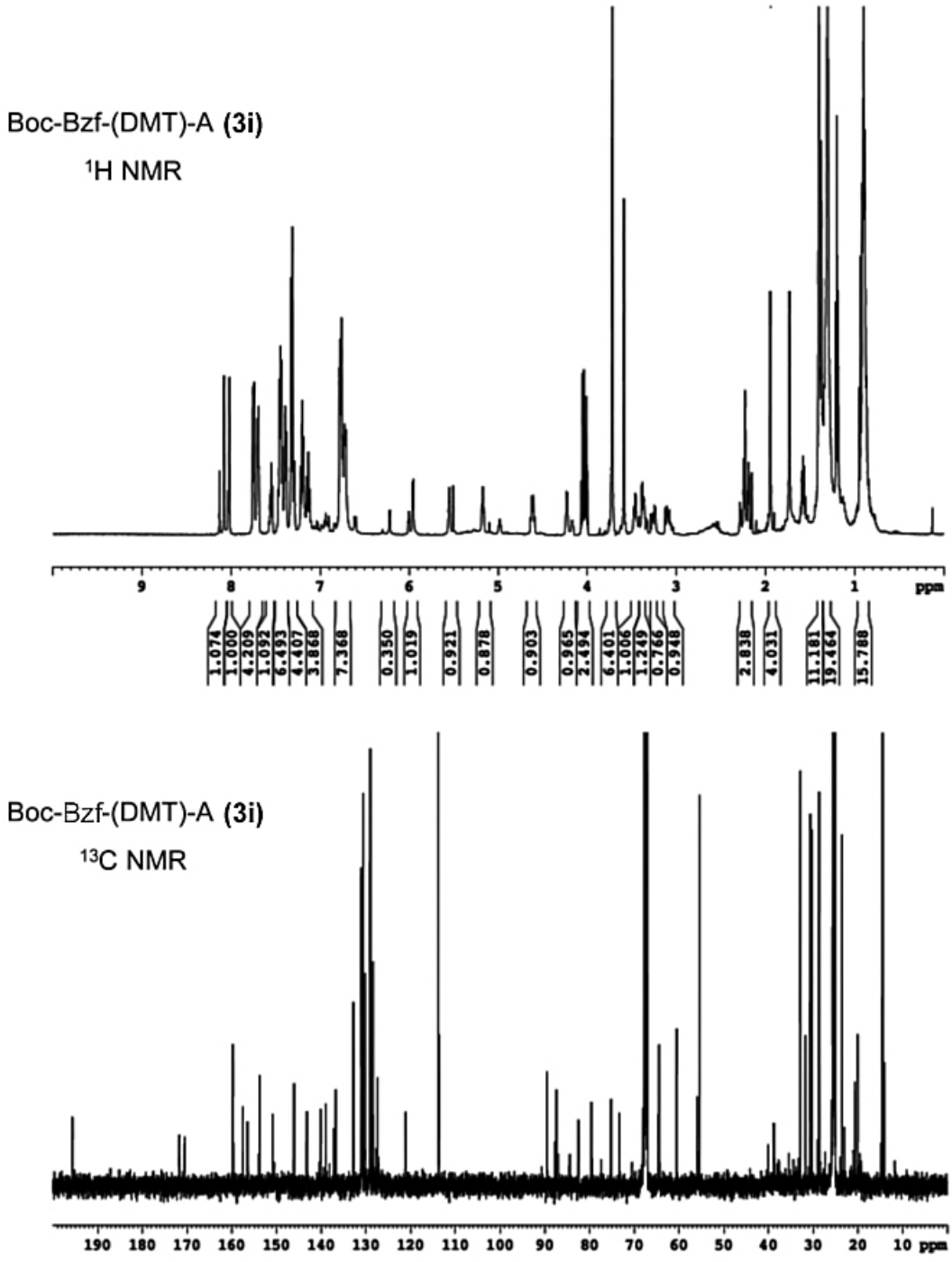


Figure 2-12. ^{1}H and ^{13}C NMR Characterization of Boc-Bzf-(DMT)-A (3i)

2.3.4 – HPLC Analysis of Donor Deprotection

Trifluoroacetic acid (1 mL) and TIPSH (10 mg, 13 µl, 65 µmol) were added to **3a** (10 mg, 13 µmol) and stirred for 12 h. Solvent was removed under reduced pressure followed by precipitation with diethyl ether. HPLC analysis of the deprotection reaction is shown in Figure 2-13. The assignment of the 2' and 3' acetylated forms of Phe-A is based on the known thermodynamic preference for 3' acylation of adenosine.[140] The HPLC method (Solvents A and B defined in main text) had the following solvent gradient (Gradient 3): 0 min 2 % B, 60 min 100 % B, 65 min 1 % B. Monitored at 215 nm and 260 nm.

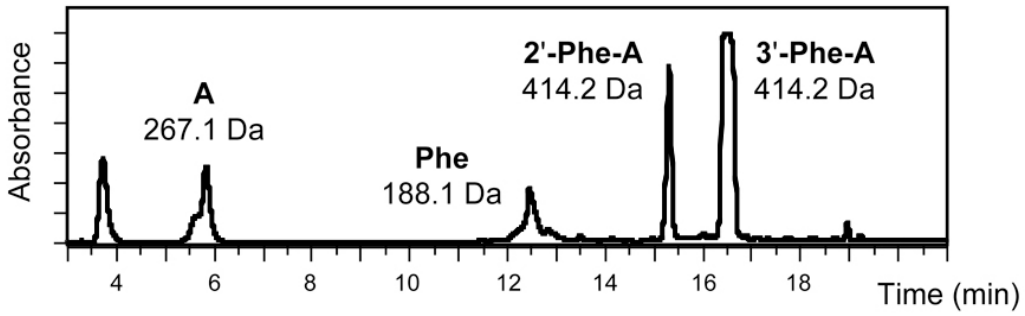


Figure 2-13. HPLC Analysis of Donor 4a Purity After TFA Deprotection

2.3.5 – HPLC Analysis of Donor Hydrolysis

The amino acid analog donor (**4a–h**) was suspected to be hydrolytically unstable and was tested in mock LysAlaAcm reactions without enzyme present and then analyzed by HPLC. The mock LysAlaAcm reactions were performed as described in section 2.3.7; however, the AaT solution was replaced with water. Phe-A solutions were analyzed at 30 min, 1 h, and 4 h. Immediately following the reaction, the sample was diluted to 1200 µL and injected onto the C18 HPLC column using Gradient 1. As seen in Figure 2-14, Phe-A is completely hydrolyzed after 4 h. Additionally, 1 mM A and 1 mM Phe were

injected on the HPLC to serve as standards to verify retention times. Note: The relative concentrations of 2' and 3' Phe-A would be different in an actual reaction where the 3' Phe-A is preferentially consumed by AaT in the transfer reaction.[43] The change in 2':3' ratio between 0 h and 30 min probably reflects the higher pH of the 30 min data.[140]

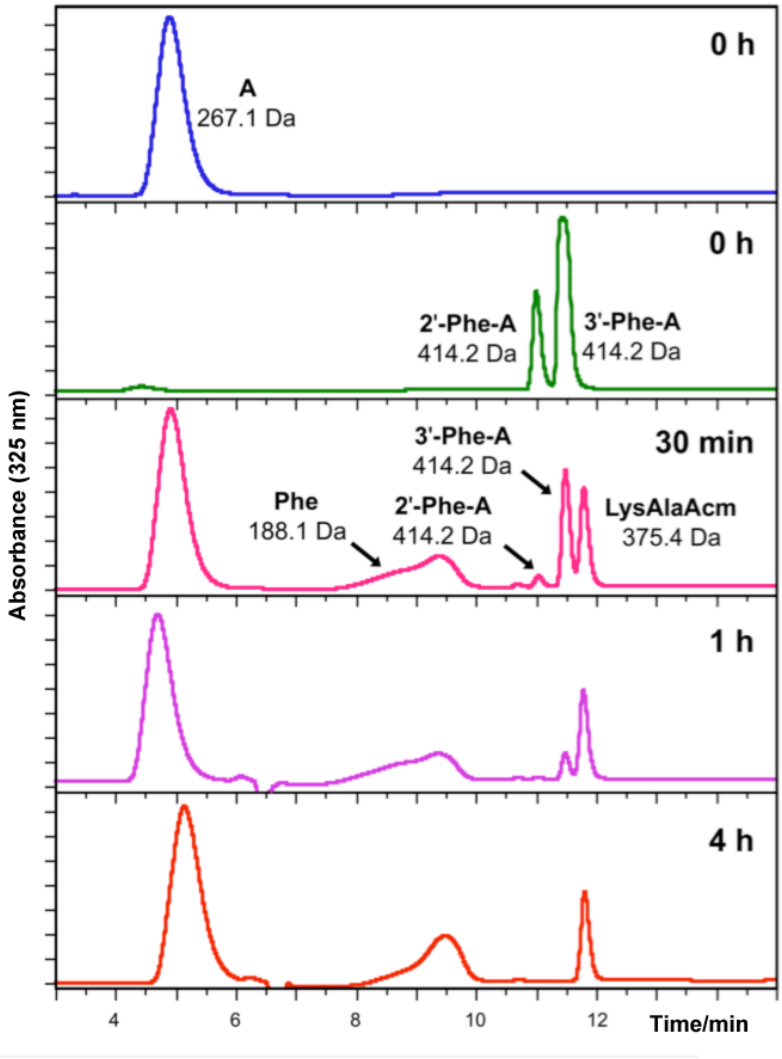


Figure 2-14. HPLC Analysis of Donor 4a Hydrolysis in Mock Transferase Reactions
Slight differences in retention time were observed, compound identities confirmed by MALDI MS.

2.3.6 – Aminoacyl Transferase Expression and Purification

E. coli AaT was expressed from the pEG6 plasmid in *E. coli* BL21-Gold (DE3) cells using a procedure adapted from Graciet *et al* (Figure 2-15).[50] *E. coli* were grown in a primary culture of 5 mL LB at 37 °C to OD_{600} of 0.5 and then were rediluted into a secondary culture of 500 mL LB and grown to OD_{600} of 0.6. AaT expression was induced using 0.1 mM isopropyl β-D-thiogalactoside and cells were grown at 25 °C for ~16 h. Cells were pelleted at 6,000 RPM using a GS3 rotor and Sorvall RC-5 centrifuge. Cell pellets were resuspended in the Ni-NTA binding buffer (50 mM Tris, 10 mM imidazole, 300 mM KCl, and 5 mM β-mercaptoethanol, pH 8.0) and included protease inhibitor cocktail, 1 mM PMSF, and 10 units/mL DNAse1–Grade II. Following resuspension, the cells were lysed using sonication. Soluble proteins were collected *via* centrifugation at 13, 200 RPM for 15 min. Collected soluble protein was gently shaken for 1 h at ambient temperature with Ni-NTA resin. The resin was prepared by rinsing with Ni-NTA binding buffer and then washed with four volumes of Ni-NTA wash buffer (50 mM Tris, 50 mM imidazole, 300 mM KCl, and 5 mM β-mercaptoethanol, pH 8.0). The proteins were eluted with elution buffer (50 mM Tris, 250 mM imidazole, 300 mM KCl, and 5 mM β-mercaptoethanol, pH 8.0). Pure elution fractions of *E. coli* AaT were dialyzed overnight in transferase buffer (50 mM Tris, 30 % glycerol, 120 mM $(NH_4)_2SO_4$, 5 mM β-mercaptoethanol, pH 8.0) at 4 °C. The dialyzed enzymes were stored at – 80°C. Protein concentrations were determined using the Bradford assay and a bovine serum albumin standard curve according to the manufacturer's instructions.[141]

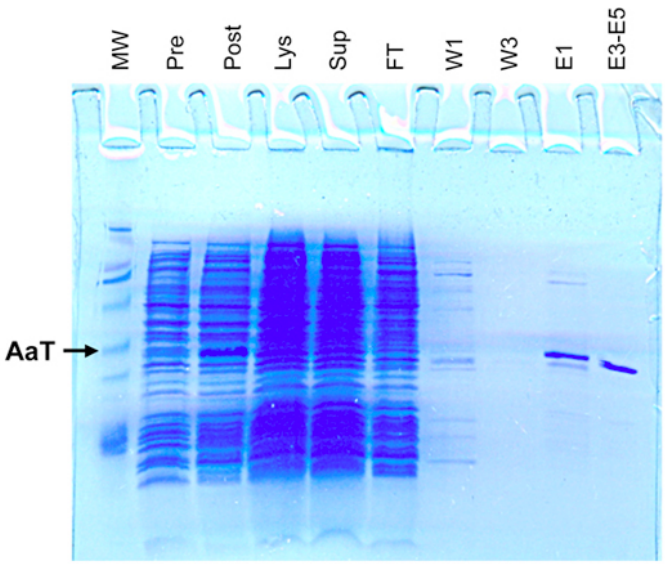


Figure 2-15. SDS PAGE Gel Analysis of AaT Expression and Purification
Lanes (left to right): 1) Molecular weight markers (Masses in kDa: 16, 25, 32, 47, 80, 100, 210); 2) Pre-induction; 3) Post-induction; 4) Crude cell lysate; 5) Supernatant after centrifugation; 6) Ni-NTA column flow-through; 7) Column wash 1; 8) Column wash 3; 9) Elution fraction 1; 10) Combined elution fractions 3-5 after overnight dialysis.

2.3.7 – LysAlaAcm Ligation Assay

Each ligation reaction, 125 µL total volume, contained the following reagents: aminoacyl adenosine donor (1 mM), recombinant His_{10}-tagged *E. coli* aminoacyl transferase (2.26 µM), and LysAlaAcm (100 µM) in the AaT Ligation Buffer (50 mM HEPES pH 8.0, 150 mM KCl, 10 mM $MgCl_2$). The reaction mixtures were incubated at 37 °C for four hours and quenched with 1 % acetic acid. The proteins were extracted from the reactions *via* acetone precipitation. The reactions were precipitated using 4X reaction volume of acetone and cooled at -20 °C for 1 h. Next, the reactions were centrifuged at 13, 200 rpm at 4 °C for 20 min to separate the reaction from precipitated protein. The supernatant was transferred to fresh 1.5 mL centrifuge tubes and allowed for acetone evaporation overnight at room temperature. After acetone evaporation, the

supernatant was dried in a Speedvac (Savant, Thermo Scientific, Fisher Inc.) for 30 min to remove residual acetone. The resulting reaction volume was dissolved up to 1.2 mL using Milli-Q water and analyzed by HPLC (gradient below) to determine ligation yield by integration of separated peak intensities monitored at 325 nm. Collected HPLC fractions were characterized through MALDI MS analysis. Reactions were performed with 6 trials on at least two protein preparations for all successful reactions, 3 trials for failed reactions. Reactions using Phe-pdCpA as donor were carried out in an identical fashion.[142]

2.3.8 – AaT Ligation Reaction Analysis by HPLC

All HPLC analyses of LysAlaAcm ligations were monitored as described in the main text (Gradient 1). Peptide retention times monitored at 325 nm (Acm absorption) and MALDI MS analyses confirming peptide identity are in Table 2-2.

Table 2-2. HPLC Analysis of LysAlaAcm Transfer Reactions

Product Peptide	Retention Time (min)	$M+H^+$	
		Calc'd	Obs.
LysAlaAcm	11.8	375.2	374.9
PheLysAlaAcm	13.0	522.3	522.1
LeuLysAlaAcm	12.7	488.3	487.3
AzfLysAlaAcm	13.6	563.3 (- N_2 537.3)	563.3 (- N_2 537.3)
NapLysAlaAcm	14.1	572.3	572.1
MefLysAlaAcm	13.1	536.3	536.3
AcfLysAlaAcm	NA	NA	NA
N_3fLysAlaAcm	NA	NA	NA
McmLysAlaAcm	NA	NA	NA
BzfLysAlaAcm	14.6	626.3	626.1

2.3.9 – A/AMP/ATP/pdCpA Inhibition Studies

Each ligation reaction was performed as described above. The addition of four types of adenosyl compounds (adenosine, AMP, ATP, or pdCpA) were monitored individually in ligation reactions with Phe donor **4a**. Adenosyl compound concentrations tested were 1 mM, 2.5 mM, and 5 mM HPLC analysis was used to determine inhibition of ligation by integration of HPLC reagent and product peaks. The LysAlaAcm ligation assay was used to analyze the kinetics of AaT with substrate **4a**. The ligation assay was modified for ease of analysis as follows: the total reaction volume was scaled up to 590 μL and all reagents were maintained at the same concentrations as mentioned above. Five concentrations of **4a** were monitored for a total reaction time of 30 min, with concentrations ranging from 0.05 to 1 mM. Each 30 min kinetic experiment was done in triplicate with a total of fourteen 40 μL aliquots for all substrate concentrations. The synthesis of pdCpA was carried out essentially as previously described.

2.3.10 – AaT Ligation Reaction Analysis by HPLC

All HPLC analyses of LysAlaAcm ligations were monitored on an Agilent HPLC using a Waters C18 column (Milford, MA). The solvents used for peptide purification were the following: 0.1 % trifluoroacetic acid in water (Solvent A) and 0.1 % trifluoroacetic acid in acetonitrile (Solvent B). The HPLC method had the following solvent gradient (Gradient 1): 0 min 1 % B, 5 min 1 % B, 10 min 30 % B, 15 min 40 % B, 20 min 100 % B, 25 min 100 % B, 27 min 1 % B, 30 min 1 % B. Peptides were monitored at two absorption wavelengths during HPLC analysis, 215 nm for peptide absorption, and 325 nm for Acm absorption. MALDI MS analyses confirmed identity of

peptide, Xaa-A donor, and ligated products.

2.3.11 – Fluorometric Analysis of Transferase Experiments

Real time fluorescent monitoring of Nap addition to LysAlaAcm was carried out using reaction solutions prepared on the 500 μL scale in stirred quartz cuvettes with 1.00 cm path lengths. Reagent concentrations were as described above, except that donor **4d** was withheld until the cuvette was placed in the fluorometer and temperature equilibrated. Reactions were then initiated by addition of 20 μL of 25 mM Nap-A (**4d**) stock or water (for negative controls). Fluorescence emission at 390 nm was monitored after excitation at 325 nm. Prior calibration had shown that NapLysAlaAcm (**5d**) concentrations could be estimated from fluorescence intensity readings (F_{5d}) using the equation $[5d] = (1 - (F_{5d}/ F_{Control}))1.39[LysAlaAcm]$, where $F_{Control}$ is the fluorescence reading of the reaction with water added instead of **4d.** Measurements were taken on the Cary Eclipse fluorometer in the "Kinetics" mode at 37 °C, acquiring data every 15 s. The excitation and emission slit widths were 5 nm and the averaging time was 1 s.

2.3.12 – Calibration of NapLysAlaAcm Fluorometric Transfer Analysis

In order to monitor conversion of LysAlaAcm to NapLysAlaAcm (**5d**) in real time, we designed a fluorometric assay based on intramolecular quenching of aminocoumarin fluorescence by the naphthyl ring. Equimolar solutions of LysAlaAcm and NapLysAlaAcm, as well as mixtures of the two solutions, were prepared and their fluorescence compared (Figure 2-16A). The fluorescence emission of NapLysAlaAcm at 390 nm (λ_{ex} = 325 nm) was 72 % lower than LysAlaAcm, and no non-linear effects were

exhibited in the mixtures (Figure 2-16B). Therefore, the mole fraction of NapLysAlaAcm could be linearly related to the measured fluorescence of a reaction solution (F_{5d}), and the concentration of **5d** determined as (1 - ($F_{5d}/F_{Control}$))•1.39•[LysAlaAcm], where $F_{Control}$ is the fluorescence reading of the reaction with water added instead of donor **4d**. This calibration was used to determine the time-dependent concentration of the NapLysAlaAcm product as described in the main text.

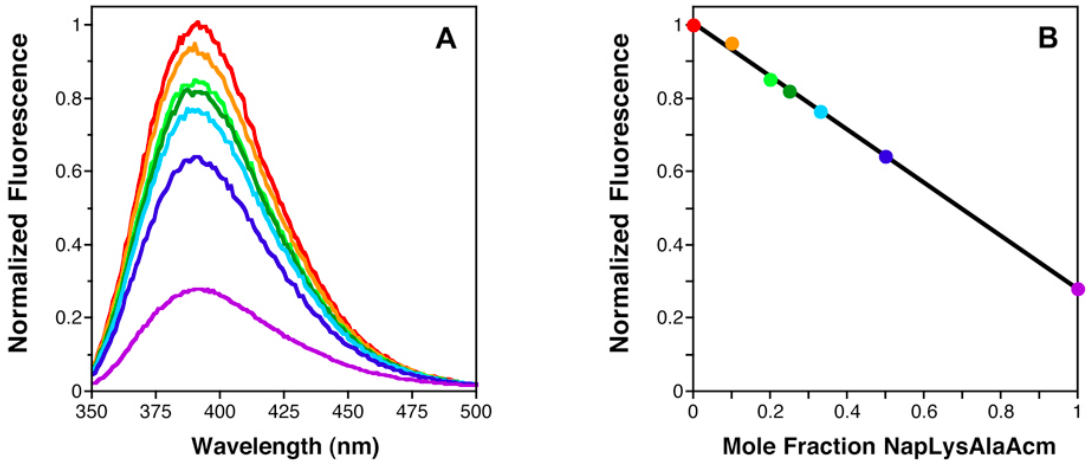
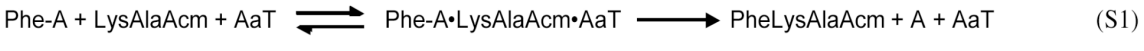


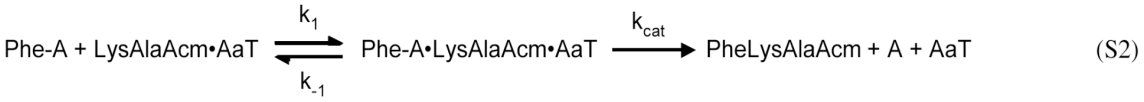
Figure 2-16. Fluorescence Emission of NapLysAlaAcm Solutions
(A) Fluorescence emission of mixtures of LysAlaAcm and NapLysAlaAcm (5d); χ5d = 0.00 (red), 0.10 (orange), 0.20 (light green), 0.25 (green), 0.33 (light blue), 0.50 (blue), 1.00 (purple)(B) Linear fit of normalized fluorescence vs. mole fraction NapLysAlaAcm (y = 1 – 0.72x, R = 0.999)

2.3.13 – Michaelis-Menten Kinetic Analysis of Phe Transfer

The reaction scheme for the catalysis of N-terminal aminoacylation by AaT is shown in Equation S1. Previously, it has been shown that the sequential bisubstrate AaT reaction can be treated as pseudo-first order when carried out with saturating concentrations of one substrate (acceptor peptide).[64] Therefore, we characterized the reaction in terms of the Michaelis-Menten kinetic scheme in Equation S2.



(S1)



(S2)

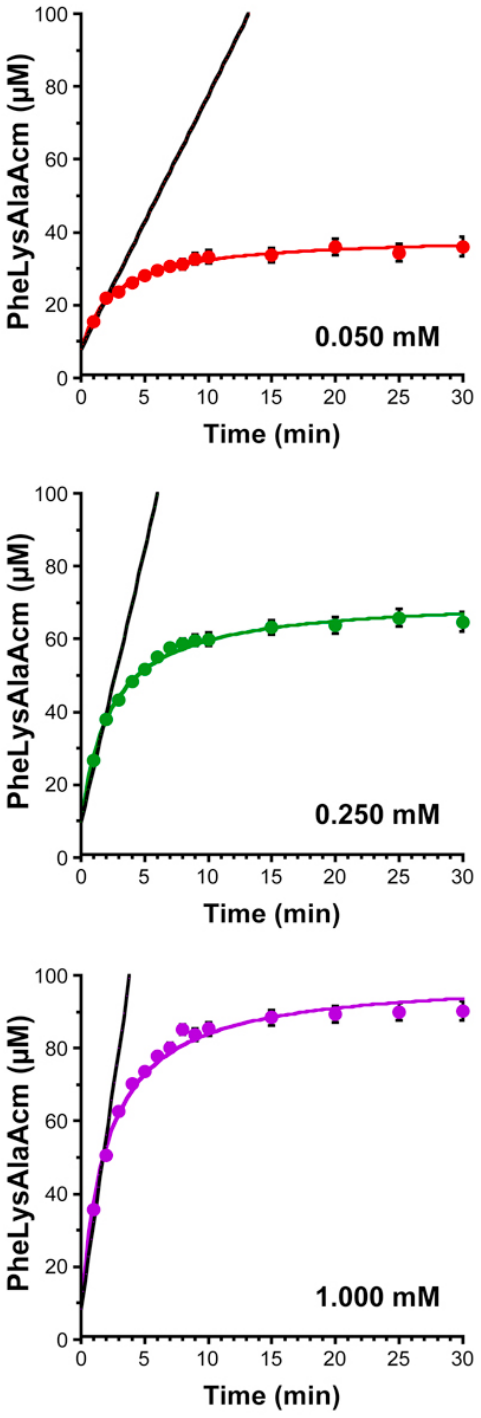
To determine initial velocities, we fit the primary kinetic data to the hyperbolic Equation S3, which, though phenomenological, produces a high quality fit ($R > 0.99$ for all 5 data sets). Data from the first two minutes were fit to a linear expression (Eqn. S4). We employ a non-zero intercept for improved quality of fit; this can be justified in terms of a rapid burst phase not captured in the 1 min resolution of this assay. Note that this burst phase seems to be extremely rapid, even the 15 s intervals of our direct fluorometric assay are not sufficient to capture this phase (See Figure 3-3A). The results of these linear fits are shown in Figure 2-17.

$$[\text{PheLysAlaAcm}] = c_1 t/(c_2 + t) \quad \text{or} \quad [\mathbf{5a}] = c_1 t/(c_2 + t) \tag{S3}$$

$$[\text{PheLysAlaAcm}] = [\text{PheLysAlaAcm}]_0 + V_0 t \quad \text{or} \quad [\mathbf{5a}] = [\mathbf{5a}]_0 + V_0 t \tag{S4}$$

The initial velocities were then plotted as a function of Phe-A donor concentration and fit to Equation S5 to obtain k_{cat} and K_M values, where $K_M = (k_{-1} + k_{cat})/k_1$ (See Figure 3-3B).

$$V_0 = k_{cat}[\text{AaT}][\text{Phe-A}]/(K_M + [\text{Phe-A}]) \quad \text{or} \quad V_0 = k_{cat}[\text{AaT}][\mathbf{4a}]/(K_M + [\mathbf{4a}]) \tag{S5}$$



[**4a**] (mM)	V_0 (μM•min^{-1})	[**5a**]$_0$ (μM)
0.050	7.34 ± 0.46	7.90 ± 1.39
0.125	11.08 ± 0.86	11.75 ± 3.39
0.250	14.17 ± 0.54	9.75 ± 2.11
0.500	17.00 ± 0.82	13.24 ± 2.83
1.000	20.70 ± 0.85	8.43 ± 4.27

Figure 2-17. Phe Transfer Reaction Kinetics
Product concentration as a function of time determined by integration of HPLC traces. Data are an average of 3 trials. Bars indicate standard error. Black lines indicate best fit to early timepoints using Eqn. S4. Bottom Right: Fits to kinetic data according to Eqn. S4.

2.3.14 – α-Casein N-Terminal Modification

α-Casein (4.8 mg) was modified with Azf-A (4c) in a reaction volume of 1 mL in modified AaT buffer (50 mM HEPES pH 8.0, 150 mM KCl, 10 mM $MgCl_2$) and AaT (0.05 mg). The reaction mixture was incubated at 37 °C for 12 h, AaT was removed using Ni^{2+} resin (Ni-NTA Superflow, Qiagen), and then buffer exchanged four times into phosphate buffered saline (PBS, 12 mM NaH_2PO_4, 50 mM NaCl, 4.7 mM KCl, pH 8.0) by Spectra/Por 1 dialysis tubing (Spectrum Laboratories; Rancho Dominguez, CA). Azf-modified α-casein was used directly after buffer exchange into PBS. Aliquots (1 nmol) of α-casein were diluted into PBS (22 μL) containing either fluorescein-alkyne (FlAlk, 91 μM) or propargylamine (Alk, 227 μM) for copper-catalyzed azide-alkyne cycloaddition (CuAAC). The CuAAC reaction was initiated by addition of $CuSO_4$ (100 μM), tris-(3-hydroxypropyltriazolylmethyl) amine (THPTA, 500 μM), and sodium ascorbate (5 mM). For control reactions, equivalent volumes were replaced with either PBS or water. The reaction mixtures were incubated at 4 °C for 2 h. Reactions were boiled with gel loading dye LDS (Pierce; Rockford, IL) for 10 min at 95 °C and analyzed by SDS-PAGE. Fluorescence images were obtained with a Geldoc (BioRAD; Hercules, CA) using an excitation wavelength of 302 nm for detection of fluorescein. α-Casein was also directly visualized by Coomassie Brilliant Blue staining. Section 2.3.14 was performed by John Warner.

2.3.15 - Edman Degradation Analysis of α-Casein Modification

Phe was ligated to α-casein using **4a** as described in the main text with the following modifications. The substrate (1 mM) was 10 times more concentrated, and the

total volume was 4 times greater. This reaction was run overnight; at the 4 h timepoint, an additional 50 µL dose of 25 mM **4a** donor was added to the reaction. The His_{10}-tagged AaT was separated from the reaction *via* nickel bead purification: 100 µL Ni-NTA resin was added at a ratio of 25 µL per 12.5 µM AaT, the beads were shaken with the reaction for 2 hours, then separated *via* centrifugation at 13, 200 RPM for 2 min. The supernatant was removed, diluted to 1 mL with Milli-Q water, and dialyzed against 1X PBS (Hyclone, Fisher) overnight at 4 °C to remove any residual **4a**. The protein was electroblotted onto PVDF membrane and sent to Johns Hopkins University Synthesis and Sequencing Facility (Baltimore, MD) for Edman Degradation analysis (Figure 2-18).

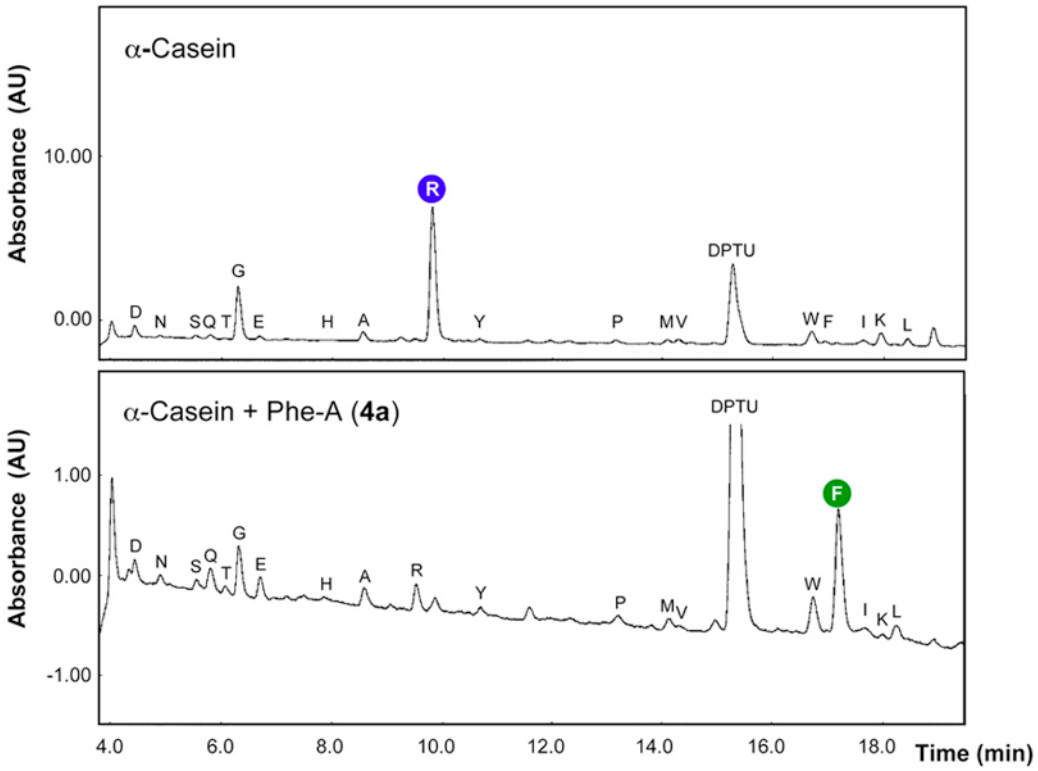


Figure 2-18. Edman Degradation Analysis of α-Casein Modification
Top: Unmodified α-casein. R peak area is 196 counts of total, F peak area is 44 counts. Bottom: Phe-modified α-casein. F peak area is 97 counts. R peak area is 53 counts.

2.3.16 – Tryptic Digest Analysis of α-Casein Modification

α-Casein-Azf (33 μg), or α-casein (33 μg), was digested with sequencing grade modified Trypsin (0.6 μg) in 27 μL of 25 mM ammonium bicarbonate pH 7.5 (freshly prepared). Digestions were carried out at 37 °C for 14 h. Trypsin digest aliquots (1 μL) were combined with α-cyano-4-hydroxycinnamic acid (1 μL of a saturated solution in 1:1 H_2O/CH_3CN with 1 % TFA) and analyzed by MALDI MS. Section 2.3.16 was performed by John Warner.

2.3.17 – α-Casein Labeling in Cleared Cell Lysate

AaT was expressed in *E. coli* BL21-Gold (DE3) cells as previously described. A cleared cell lysate was obtained by centrifugation following cell lysis using sonication. Proteins were modified in a final reaction volume of 110 μL with Azf-A (4c, 3 mM) using AaT (25 μL) from cleared lysate in AaT buffer (50mM HEPES pH 8.0, 150mM KCl, 10mM $MgCl_2$). α-Casein (0.012 mg or 0.12 mg) was added to the reaction and incubated at 37 °C for 2 h, after 1 h, an additional 0.3 μmole of 4c was added. For control reactions, equivalent volumes were replaced with water. Excess 4c was removed by buffer exchange four times against phosphate buffered saline (PBS, 12mM NaH_2PO_4, 50mM NaCl, 4.7mM KCl, pH 8.0) by Spectra/Por 1 dialysis tubing. Aliquots (18 μL) of **4c** modified lysates were diluted to a final volume of 22 μL using PBS containing FlAlk (182 μM) for copper catalyzed "click" reactions. The "click" reaction was initiated by addition of $CuSO_4$ (100 μM), THPTA (500 μM), and sodium ascorbate (5 mM). The reaction mixtures analyzed by PAGE gel using Coomassie staining and fluorescence as above (Figure 2-19). Section 2.3.17 was performed by John Warner.

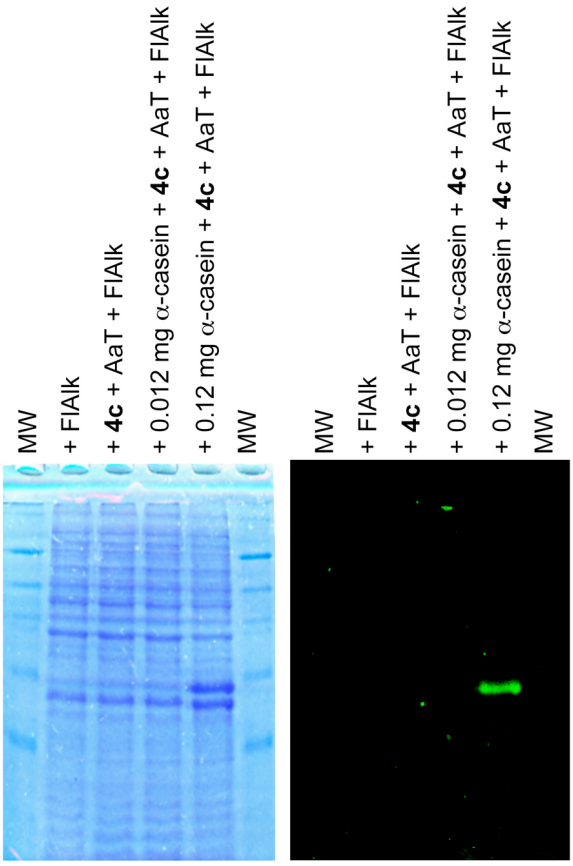


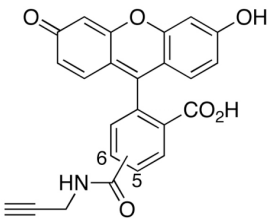
Figure 2-19. SDS PAGE Gel Analysis of α-Casein Modification in Cleared E. coli lysate
Left: Coomassie stained gel. Right: Fluorescence image from 302 nm excitation. Lanes (left to right): 1) Molecular weight markers (Masses in kDa: 17, 25, 30, 46, 58, 80, 175); 2) Cleared lysate reacted with FlAlk; 3) Cleared lysate modified with **4c** then reacted with FlAlk; 4) Cleared lysate with α-casein (0.012 mg) modified with **4c** then reacted with FlAlk; 5) Cleared lysate with α-casein (0.12 mg) modified with **4c** then reacted with FlAlk; 6) Molecular weight markers.

2.3.18 – "Click" Reagent Synthesis

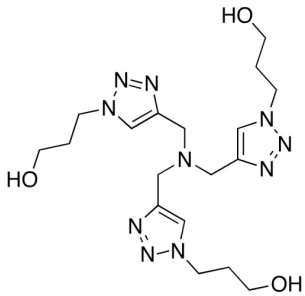
Fluorescein 5- and 6-proparglyamide (Fluorescein-alkyne, FlAlk) was synthesized as previously described (Figure 2-20).[134] Briefly, propargylamine (8.7 µL, 136 µmol) was added to a solution of 5- and 6-Carboxyfluorescein *N*-hydroxysuccinimide ester (32 mg, 68 µmol) in tetrahydrofuran (5 mL). The reaction was stirred continuously and monitored by TLC. Solvent was removed by under reduced pressure. Purification by

silica chromatography (R_f = 0.3, 20 % MeOH in CH_2Cl_2) afforded fluorescein-alkyne as a red-orange solid. Tris-(3-hydroxypropyltriazolylmethyl)amine (THPTA) was also synthesized as previously described (Figure 2-20).[143] Briefly, to a solution of 3-bromo-propanol (2 g, 14.4 mmol) in dichloromethane (10 mL) was added Ac_2O (2.94 g, 28.8 mmol) and NEt_3 (2.91 g, 28.8 mmol). The reaction was stirred continuously and monitored by TLC. An aqueous solution of $NaHCO_3$ was added and the phases were separated. The organic layer was washed once more with $NaHCO_3$ and twice with brine. Solvent was removed under reduced pressure and 3-bromopropyl acetate was afforded as a colorless oil. To a solution of 3-bromopropyl acetate (1.0 g, 6.0 mmol) in water (10 mL) was added sodium azide (780 mg, 12.1 mmol). The solution was stirred at 90 °C overnight. The solution was extracted three times with 15 mL CH_2Cl_2 and dried with $MgSO_4$. Solvent was removed under reduced pressure to afford 3-azidopropyl acetate as a pale yellow oil. To a solution of tripropargylamine (12.3 mg, 0.941 mmol) in tetrahydrofuran (6.5 mL) was added 3-azidopropyl acetate (0.672 g, 4.71 mmol) and Cu(I) acetate (5.8 mg, 5 mol %). The solution was refluxed overnight under inert atmosphere. Solvent was removed under reduced pressure. Silica chromatography purification (R_f = 0.5, 5-10 % MeOH in CH_2Cl_2) afforded acetyl-protected THPTA as a yellow oil. A solution of acetyl-protected THPTA (397 mg, 0.711 mmol) in 2.0 M ammonia in MeOH (10 mL) was continuously stirred overnight at 40 °C. Solvent was removed under reduced pressure. The solid was washed and filtered four times with acetonitrile and dried under reduced pressure to afford THPTA as a white solid. For both FlAlk (25 % overall yield) and THPTA (98 % overall yield), intermediate and product

identities were confirmed by comparison of ^{1}H NMR and LRMS (ESI) MS data to previous reports.[134, 143] Section 2.3.18 was performed by John Warner.



Fluorescein 5- and 6-Proparglyamide
FlAlk



Tris(hydroxypropyltriazoylmethyl)amine
THPTA

Figure 2-20. Click Chemistry Reagents

CHAPTER 3

NATIVE CHEMICAL LIGATIONS IN PEPTIDES AND PROTEINS USING PROTECTED DISULFIDE AMINOACYL TRANSFERASE ANALOGS

Section 3.1 - Introduction

Protein engineering is an expansive field that has a wide-ranging definition from designing and creating new proteins to altering proteins to have a desired function.[16] Despite the breadth of the field, there is an overarching goal to analyze and understand protein structure and function. One recent method has revolutionized the field of protein synthesis, native chemical ligation (NCL). NCL component groups include a C-terminal peptidyl thioester and a N-terminal cysteine peptide that react to form an amide bond while retaining the cysteine residue at the ligation site.[9] In order to minimize the need for peptide synthesis, the NCL reaction can be used with expressed protein fragments and is termed expressed protein ligation (EPL).[144] The EPL reaction can be implemented in two methods that vary in the degree of chemical synthesis required. One method expresses an N-terminal cysteine protein fragment in *E. coli* cells and utilizes chemical synthesis to create a small thioester-bearing peptide for the ligation reaction. Another method, pioneered by Muir, expresses both protein fragment domains: an N-terminal cysteine-bearing protein fragment and a fused intein at the C-terminus. Under reducing conditions, the intein can be self-excised and form a C-terminal thioester *in situ*. Both methods, NCL and EPL, have helped to make protein synthesis an efficient process.

The major limitation to NCL and EPL is the need of a cysteine at the ligation site. Since cysteine is coded by only two codons, it will be theoretically underrepresented in proteins at a random occurrence of 3.3 % composition in proteins.[145] The percent composition of cysteine in proteins is actually lower than the random value, depending on species, 2.3 % in mammals and 0.5 % in some archeabacteria.[145] One way to bypass the constraint of low cysteine abundance in proteins is amino acid mutagenesis. However, mutagenesis to a cysteine can be disruptive to protein structure and activity. Therefore, several groups have developed methods to erase or mask the ligation site to a different residue. Several methods include alkylation of cysteine to form an amino acid analog or desulfurization of cysteine to form alanine.[146-148] Other strategies, such as one developed by Danishefsky, desulfurize synthetic cysteine analogs to a variety of other amino acids such as valine and lysine.[149-155] Post-ligation reactions, demonstrated by Dawson and Muir, utilize nitrobenzyl cysteine surrogates at ligations, which can be removed *via* photolysis.[156-157] Other ligations methods create cysteine-less sites such as "traceless" Staudinger ligations and Bode's ketoacid/alkoxyamine ligation.[30, 158-159] Additionally, homocysteine (Hcs) can be used as a cysteine surrogate for ligation and then converted to methionine by selective methylation.[160-163] All of the aforementioned strategies to mask the ligation site have one common link, the requirement of at least one synthetically incorporated reactive ligation group.

We have developed and demonstrated a method that utilizes an enzyme, aminoacyl transferase (AaT) from *E. coli*, to specifically functionalize the N-termini of peptides and proteins with a wide variety of natural and unnatural residues. Recently we

have developed aminoacyl donors that can transfer ligation handles such as Hcs, which were not previously accessible in expressed proteins. The natural function of AaT is to transfer phenylalanine, leucine, and at a lesser efficiency, methionine from an aminoacyl tRNA (Xaa-tRNA) to the N-terminus of a protein.[40, 123] In previous work, we have successfully demonstrated that AaT can transfer amino acids with large phenylalanine-like side chains such as benzoylphenylalanine and naphthylalanine.[66] We applied a similar strategy to a variety of methionine-like analogs and determined their efficiency as AaT substrates to be used to functionalize proteins with erasable Cys analogs for ligation. The applicability of AaT transfer is far reaching since it recognizes N-terminal Arg and Lys residues without influence by adjacent amino acids; both arginine and lysine are significantly more abundant than cysteine at 5.5 % and 6.4 %, respectively.[40, 127, 145] We have demonstrated that AaT can transfer disulfide-protected Hcs and Cys to the N-terminus of a protein *via* amide bond formation and in conditions that maintain protein stability and functionality. Additionally, following reduction and ligation, Hcs can be effectively and selectively masked to become Met by alkylation (Figure 3-1). We have validated the utility of this approach in two ways: by synthesizing α-synuclein (αS), a protein whose aggregation contributes to Parkinson's disease pathology, and through a multi-step semi-synthesis of the well-characterized calcium-binding protein calmodulin (CaM), to demonstrate the ability to selectively incorporate an unnatural site in the middle of a protein while retaining the original primary sequence in the rest of a Cys-less protein.[164-165] This N-terminal modification technique will expand the utility of current NCL methods.

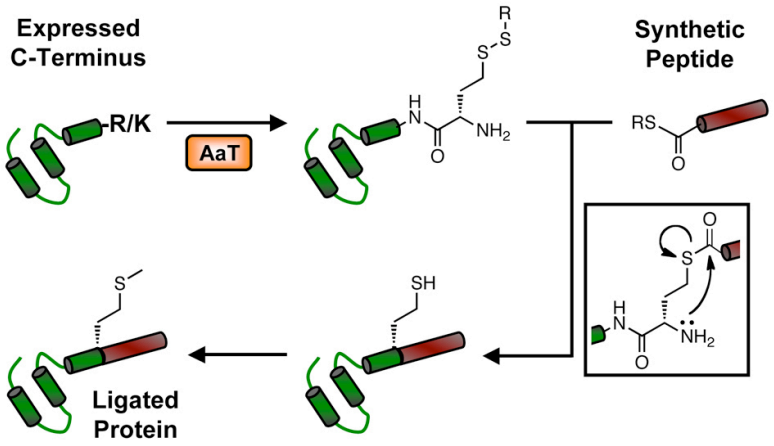


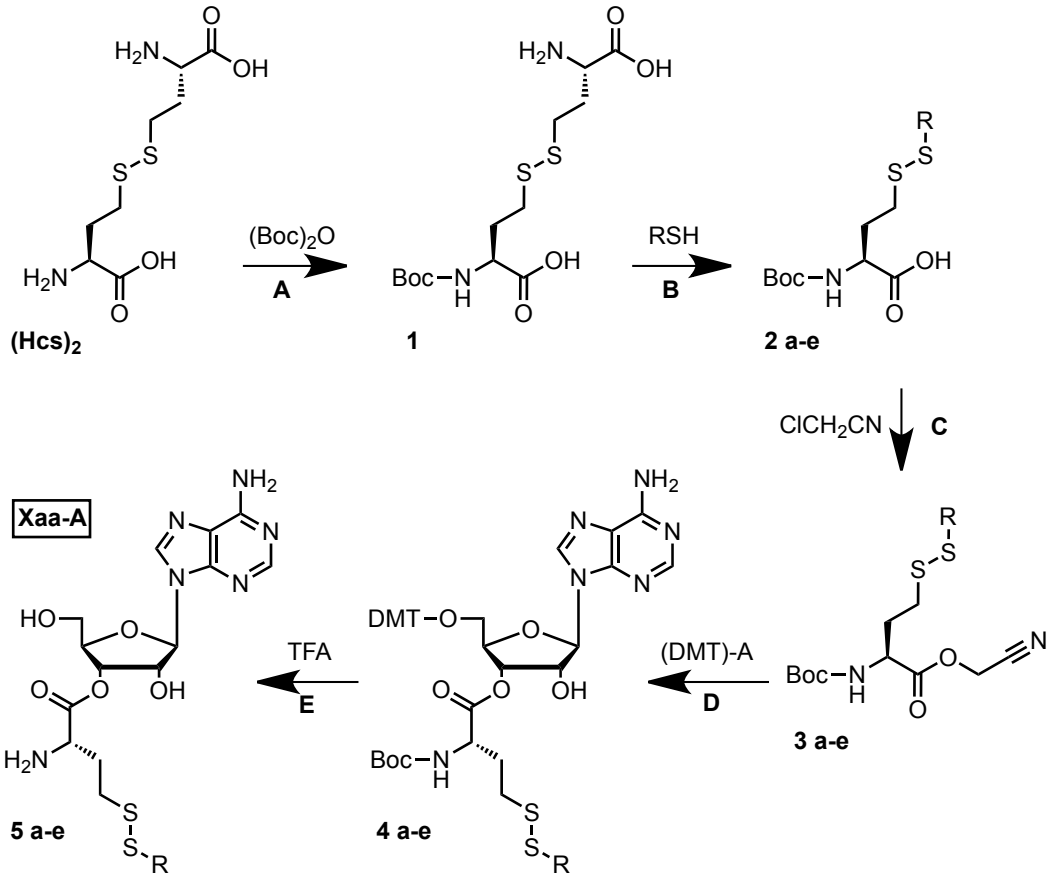
Figure 3-1. Expressed Protein Ligation at Methionine
An expressed protein bearing an N-terminal Arg or Lys (green) is functionalized with a protected Hcs analog from an aminoacyl adenosine donor, under catalysis by AaT transferase. A synthetic peptide thioester (red) is ligated to the modified protein under reducing conditions that allow deprotection of Hcs. After ligation, Hcs is converted to Met *via* methylation.

Section 3.2 - Results and Discussion

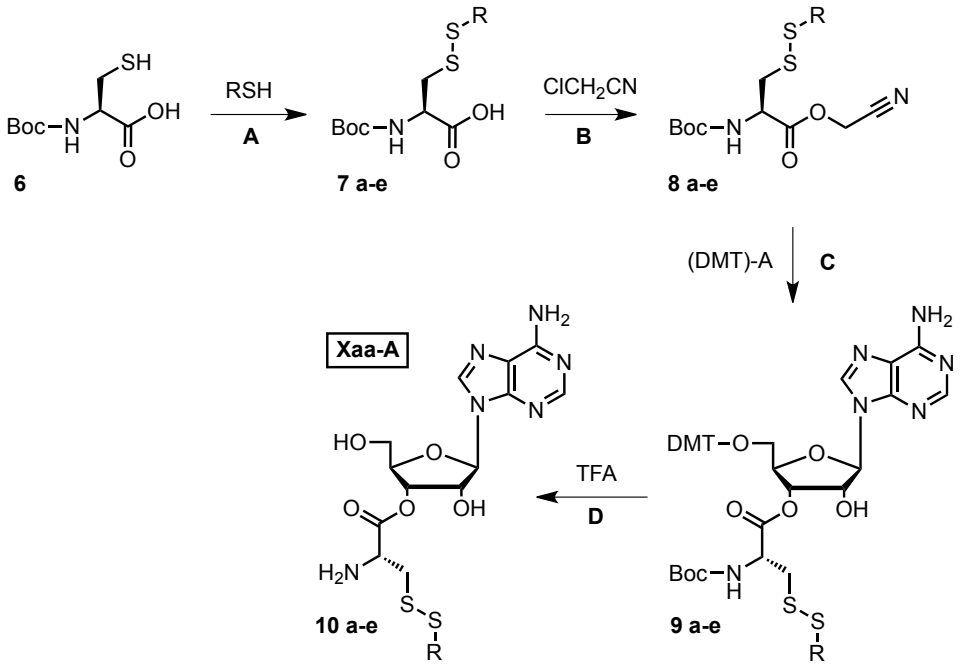
Methionine is the only sulfur-containing amino acid that is naturally transferred by AaT in *E. coli*. Although methionine is not a substrate for NCL, Hcs and other thiol- or selenol-containing Cys analogs can be used in ligation . Therefore, we investigated the possibilities of AaT ligation of Hcs and Cys (as a test for various Cys analogs such as selenocysteine) in the form of adenosyl analogs, which bypass the restrictive selectivity of an aminoacyl synthetase. Previous studies had shown that Cys-tRNA is not a substrate for AaT and our own preliminary experiments showed that Hcs was not transferred (data not shown).[58, 127] Therefore, we prepared Hcs analogs masked as aliphatic or aromatic disulfides in order to mimic the endogenous AaT substrates: Met, Leu, and Phe. Disulfide protecting groups (MeS, *i*-PrS, *t*-BuS, PhS, and CF_3CH_2S) were used so that they could be removed *in situ* under the reducing conditions of the subsequent NCL reaction.

3.2.1 - Disulfide Aminoacyl Adenosyl Donor Synthesis.

All of the aminoacyl donors for AaT were created using a similar synthetic protocol. For the synthesis of *S*-protected 3'-adenosyl homocysteine, the reaction sequence began with Boc-protection of the exposed amino groups of homocystine to form homodisulfide **1**. The second step involved disulfide exchange with a thiol containing the desired aliphatic or aromatic protecting group to form heterodisulfides **2a-e**. Next, the carboxylic acid group was activated in the form of a cyanomethyl ester (**3a-e**). In the following step, the activated carbonyl will allow for nucelophilic attack of the 2' or 3' hydroxyls of 5'-*O*-dimethoxytrityladenosine (DMT-A) to form protected aminoacyl adenosines **4a-e**. The last step of the synthetic route was deprotection of all protecting groups (Boc and DMT) with 50 % trifluoroacetic acid and 50 % tetrahydrofuran yielding the aminoacyl adenosyl analogs **5a-e**, denoted Xaa-A (Scheme 3-1). Overall yields of products **4a-e** from homocystine were ~ 40 %. To ensure high purity of final product and removal of deprotection side products, the final step was purified by HPLC. The syntheses of Cys analogs were carried out in a similar fashion (Scheme 3-2).



Scheme 3-1. Synthesis of Homocysteine (Hcs) Analogs
All homocysteine (Hcs) analogs began with the starting material homocystine, $(Hcs)_2$. The R-groups for each letter is as follows: **2a-5a** R = methyl, **2b-5b** R = isopropyl, **2c-5c** R = tertbutyl, **2d-5d** R = phenyl, and **2e-5e** R = trifluoroethyl. Reactions and conditions for step (A) 2.5 equivalents of $(Boc)_2O$, 4 equivalents of NaOH, 50/50 dioxane/water, (B) excess RSH (R = protecting group), I_2, 50/50 THF/water, (C) 1.1 equivalents of DIPEA, $ClCH_2CN$, (D) 2 – 4 equivalents of cyanomethylester, 1 equivalent of DMT-A, catalytic equivalent of TBAC, THF, (E) 4 equivalents TIPSH, 50/50 TFA/THF. Yields can be found in section 3.3 Supplementary Materials and Methods. Boc = *tert*-butyloxycarbonyl, DIPEA = *N*-diisopropylethylamine, THF = tetrahydrofuran, (DMT)-A = 5′ -*O*-dimethoxytrityl adenosine, NBu_4Ac = *N*,*N*,*N*,*N*-tetrabutylammonium acetate, TFA = trifluoroacetic acid, TIPSH = triisopropylsilane.



Scheme 3-2. Synthesis of Cysteine (Cys) Analogs
All Reagents and conditions: a) NaSMe, I_2, 1:1 H_2O/THF, 0 °C; b) $ClCH_2CN$, DIPEA, THF; c) (DMT)-A, NBu_4Ac, THF; d) TFA, TIPSH, THF. The R-groups for each letter is as follows: **7a-10a** R = methyl, **7b-10b** R = isopropyl, **7c-10c** R = tertbutyl, **7d-10d** R = phenyl, and **7e-10e** R = trifluoroethyl. Reactions and conditions for step (A) excess RSH (R = protecting group), I_2, 50/50 THF/water, (B) 1.1 equivalents of DIPEA, $ClCH_2CN$, (C) 2 – 4 equivalents of cyanomethylester, 1 equivalent of DMT-A, catalytic equivalent of TBAC, THF, (D) 4 equivalents TIPSH, 50/50 TFA/THF. Yields can be found in section 3.3 Supplementary Materials and Methods. Some analogs were commercially-bought as the Boc-protected amino acid, including Cys-*S*-t-butyl (Csb) and Cys-benzyl (Cbz). All cysteine disulfide cyanomethylesters were synthesized by Dr. Tomohiro Tanaka except for Csb and Cbz, which were synthesized by Anne Wagner and Jose Villegas. The rest of the synthetic pathway was completed by Anne Wagner.

3.2.2 - LysAlaAcm Model Ligations with Disulfide Analogs

As in Chapter 2, we used LysAlaAcm as a model peptide to test the feasibility of disulfide analogs as aminoacyl-tRNA substrate mimics for AaT. All reactions were kept standardized, as discussed in Chapter 2, with one additional change: all reactions had a total of 4 doses of 1 mM Xaa-A added at t = 0 h, 1h, 2h, and 3 h time-points of the 4 h reaction. This change significantly enhanced ligation efficiency by making more analog available throughout the 4 h reaction to compete against analog hydrolysis. We

compared yields of XaaLysAlaAcm by HPLC analysis after four hour reactions at pH 8.0 at 37 °C (Figure 3-2). Under these conditions, transfer of Phe is quantitative and Met proceeds in 75 % yield. Once we determined that AaT could transfer sulfur-containing amino acids, we tested a methionine isostere, Cys-*S*-methyl (Csm), and it produced a ligation efficiency of 21 %. While this was lower than Met, it established that we could transfer disulfide-protected analogs without substantial sidereactions. Hcs donors (**5a-e**) and Cys donors (**10a-10e**) were tested in a similar fashion (Schemes 3-1 and 3-2). For homocysteine analogs, the bulk of the protecting group was a significant factor in transfer efficiency: Hcm was transferred most efficiently from **5a**, with somewhat lower yields for Hcp and Hcb. Of the cysteine disulfide analogs, Csp had the highest transfer efficiency: Csp (**10b**) > Csb (**10c**) > Csm (**10a**). All of the aromatic protecting groups (**5d** and **10d**) prevented the ligation of their respective amino acid to LysAlaAcm, including Cbz, which was not a disulfide, indicating that issues with residue steric bulk, not reactivity of the aryl-S bond. In order to probe active site pocket polarity, ligation with a trifluoroethyl protecting group (**5e** and **10e**) was attempted without success for either homocysteine or cysteine disulfide analogs.

To ensure maintenance of the disulfide-protected analogs during the transfer reaction, we eliminated β-mercaptoethanol (BME) from the AaT purification process and reaction buffer (Figure 3-3). This small change increased the yields of HcmLysAlaAcm from 21 % to 37 % and CspLysAlaAcm from 64 % to 68 %. Once Hcm was identified as the optimal homocysteine substrate, we assessed the possibility of generating Hcm-tRNA *in situ* using a synthetase. Free Hcm was synthesized by TFA deprotection of Boc-Hcm.

Wild type *E. coli* MetRS and an $L_{13}G$ mutant (Met*RS) previously used by the Tirrell group were tested with Hcm, ATP, and *E. coli* total tRNA.[166] Met*RS was more active than MetRS, giving full transfer at ≤ 0.1 mg/mL loading. This is consistent with Met*RS structural analysis, where longer sidechains are accommodated by the expanded active site.[166]

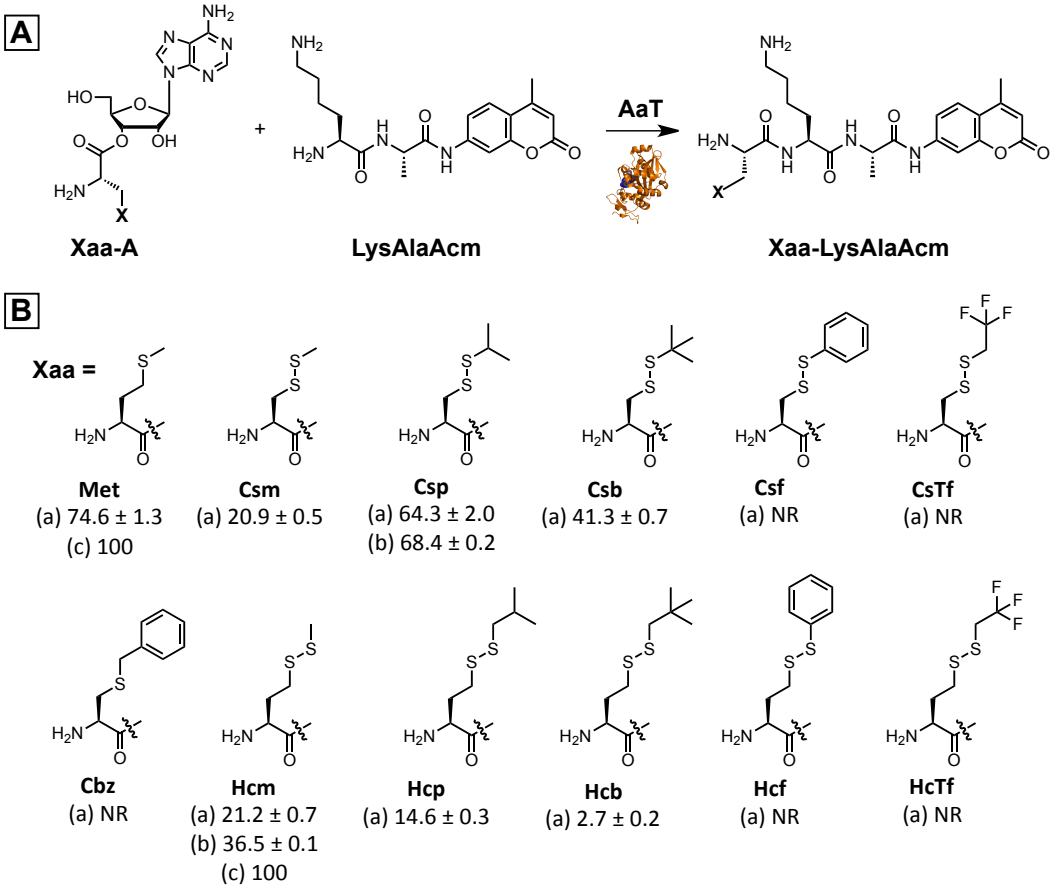


Figure 3-2. (A) Schematic Representation of Model LysAlaAcm Peptide N-Terminal Ligation Reaction Using AaT. (B) Molecular Structures of All Sulfur-Based Analogs Tested with AaT

Listed below each amino acid is its percent ligation efficiency, (a) reaction in the presence of 5 mM beta-mercaptoethanol, (b) reaction with no beta-mercaptoethanol, and (c) reaction using methionine $L_{13}G$ synthetase, free-amino acid, tRNA, and AaT. NR = no reaction.

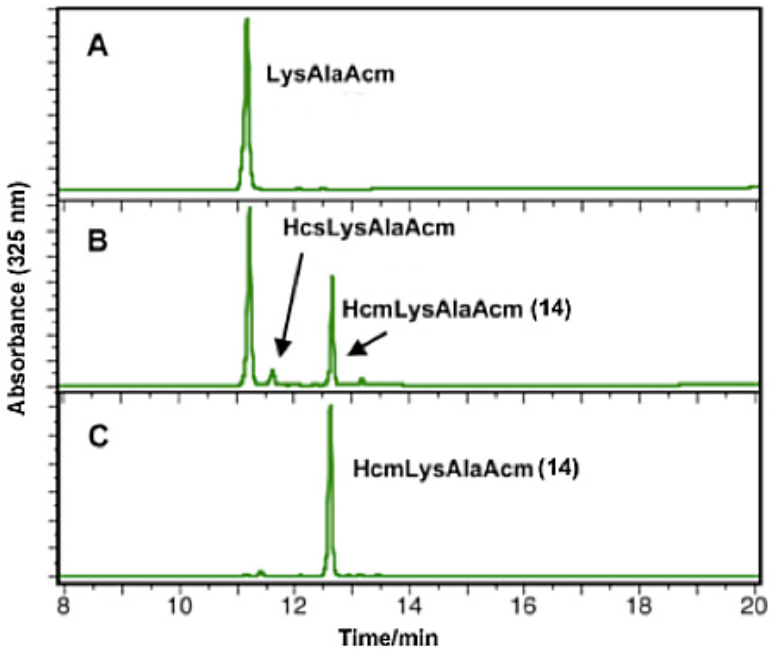


Figure 3-3. Hcm Ligation Increased Efficiency
(A) LysAlaAcm starting material, (B) Hcm ligation in presence of 5 mM BME using Hcm-A and AaT, (C) Hcm transfer using MetRS $L_{13}G$ mutant synthetase (Met*RS), tRNA ATP and AaT

3.2.3 – Hcm/Csp and Native Chemical Ligation

Our next step in analyzing our analogs' applicability to semi-synthetic techniques was to test their efficiency in NCL reactions (Figures 3-4a, 3-4b). First, we analyzed the Csp analog, which was most efficiently transferred onto our model peptide, LysAlaAcm. We ligated CspLysAlaAcm (**12**) with Ac-MetAspValPhe-SR (**11**), which is the four amino acid starting sequence of αS, and formed the NCL product (**13**) over a 24 h reaction in 100 % yield (Figure 3-5a). We then took HcmLysAlaAcm (**14**) on to a model ligation reaction with Ac-MetAspValPhe-SR (**11**). HPLC and MALDI MS analyses showed that disulfide cleavage was completed almost immediately and that the subsequent ligation reached 100 % completion after 24 h (Figure 3-5b). After ligation,

the product peptide (**15**) was treated with 1000 equiv. MeI at pH 8.6 to form **16**. After 5 min, 89 % conversion of Hcs_5 to Met_5 was observed by HPLC and MALDI MS analyses (See Materials and Methods). Unreacted **15** (11 %) could be removed after disulfide formation. Oxidation of Met_1, was observed (4 %), but no undesired alkylation of Met or Lys was seen. This is consistent with previous uses of Hcs as a ligation handle, where subsequent alkylation is surprisingly selective, given that the pK_a of Hcs (8.9) is close to the pK_a of Lys (9.5).[167]

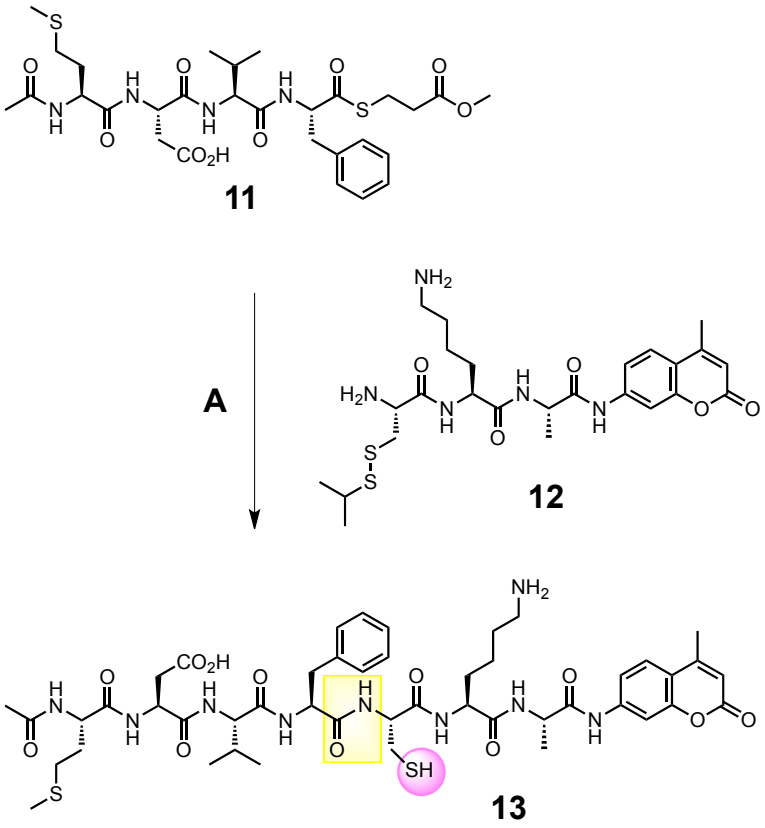


Figure 3-4a. Peptide Native Chemical Ligation and Masking Proof-of-Principle for Csp
Compound **11** is the C-terminal thioesterified first four amino acid sequence of α-synuclein. Compound **12** is Csp ligated to LysAlaAcm *via* AaT. Compound **13** is the NCL product. Reaction Conditions: **(A)** 1.0 mM (12), 3.3 mM (11), 100 mM sodium phosphate dibasic, 20 mM TCEP, 0.5 % thiophenol, 0.5 % benzylmercaptan, pH 7.2 at room temperature for 24 h, 100 % yield for **13**. The formed amide bond is highlighted in a yellow box. The resulting deprotected disulfide is highlighted in pink (Cys).

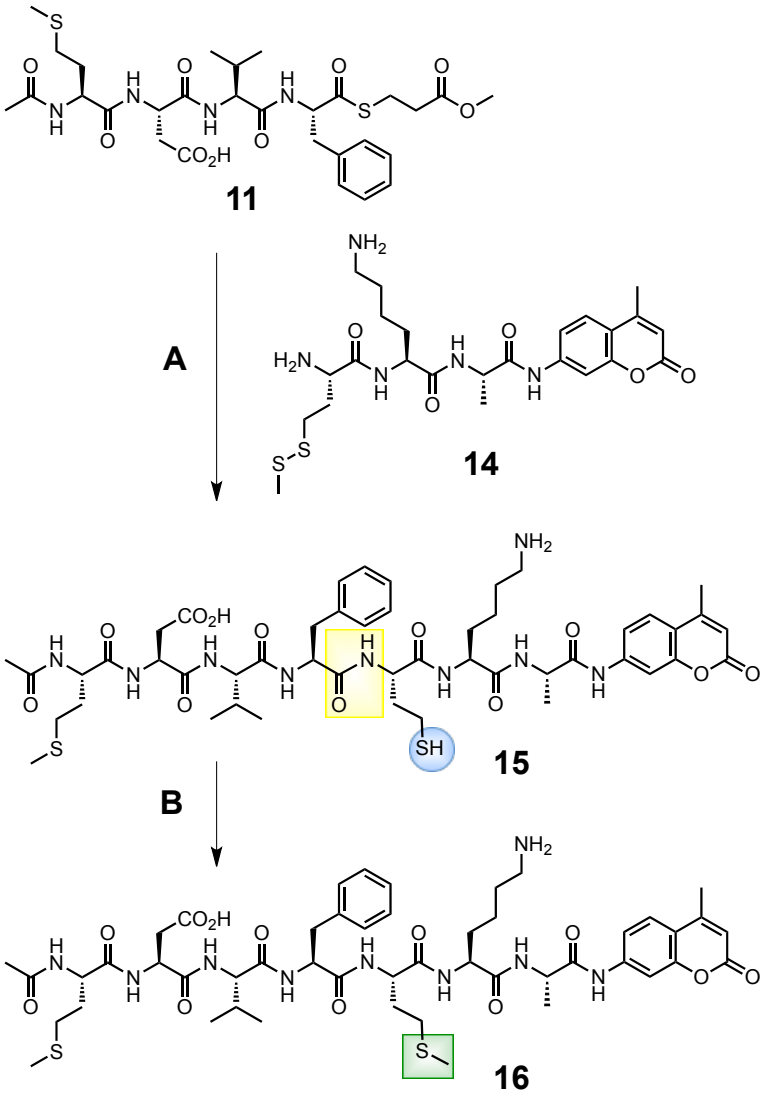


Figure 3-4b. Peptide Native Chemical Ligation and Masking Proof of Principle for Hcm
Compound **11** is the C-terminal thioesterified first four amino acid sequence of α-synuclein. Compound **14** is Hcm ligated to LysAlaAcm *via* AaT. Compound **15** is the native chemical ligation product with Hcm deprotected to Hcs. Compound **16** is the methylated masked final product, creating a traceless ligation site at the second methionine. Reaction Conditions: **(A)** 1.0 mM **13**, 3.3 mM **11**, 100 mM sodium phosphate dibasic, 20 mM TCEP, 0.5 % thiophenol, 0.5 % benzylmercaptan, pH 7.2 at room temperature for 24 h, 100 % yield for **15**; **(B)** 50 mM MeI, 100 mM sodium bicarbonate, pH 8.6 at room temperature for 5 min, 89 % yield for **16**. The formed amide bond is highlighted in a yellow box. The resulting deprotected disulfide is highlighted in blue (Hcs). The Hcs is selectively methylated, highlighted in the green box.

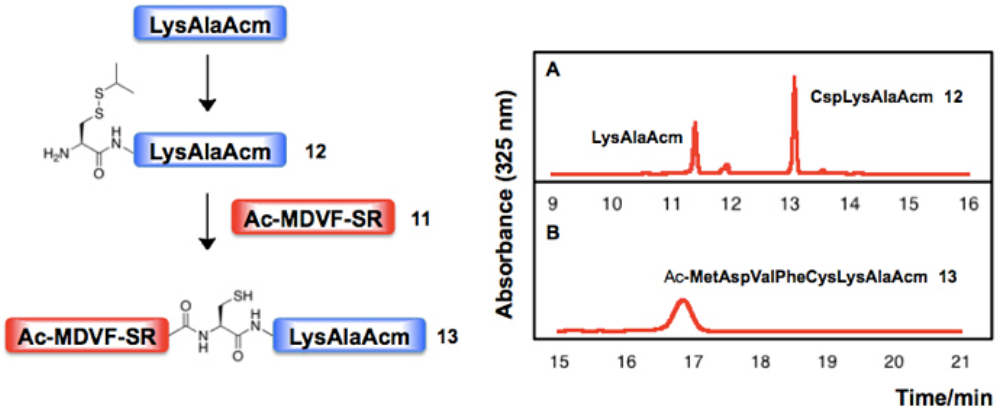


Figure 3-5a. HPLC Analysis of Peptide Ligation Reactions with Csp
(A) N-terminal Csp peptide **12** formation reaction; (B) C-terminal thioester peptide (**11**) NCL reaction with **12** after 24 h, forming product **13**.

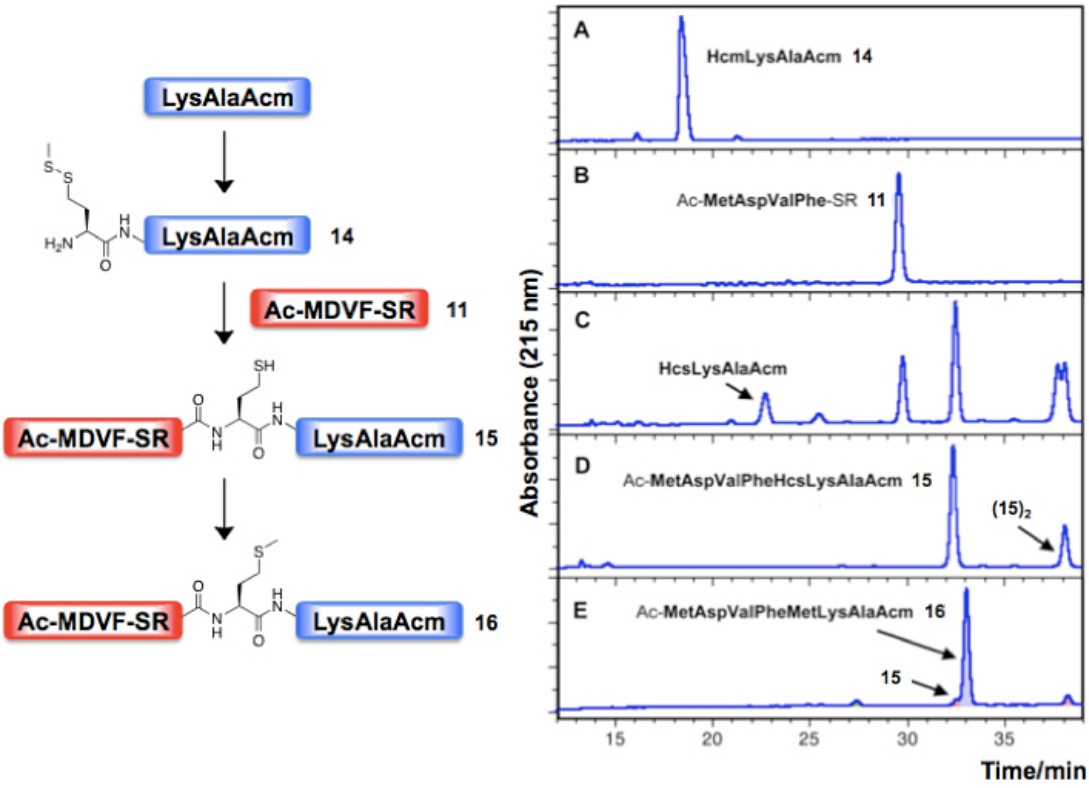


Figure 3-5b. HPLC Analysis of Peptide Ligation Reactions with Hcm
(A) N-terminal Hcm peptide **14**; (B) C-terminal thioester peptide **11**; (C) Ligation reaction after 1 h; (D) Purified ligation product **15**; (E) Methylation of **15** to form methionine in peptide **16**. **$(15)_2$** is a disulfide between the Hcs thiol of two **15** product peptides. Peak areas indicated by red, green, and blue shading were integrated in Kaleidagraph to obtain methylation yields.

After successful completion of the model peptide ligation, we wished to test the Hcm transfer/ligation strategy with a full-sized protein. One protein that we have previously studied, which has the requisite MetLys motif, is αS. Several recent studies have shown that the N-terminal sequence of αS can have important consequences for its folding and self-association.[168-172] However, these studies profoundly disagree on what those consequences are. Therefore, the preparation of N-terminally-modified αS could be useful to addressing these questions and serve as a valuable demonstration of our method.

We chose Met_5Lys_6 as the point of disconnection and prepared an $\alpha S_{6\text{-}140}$ plasmid with an N-terminal His_{10} tag and a Factor Xa proteolysis site. The precursor protein, HisTag-$\alpha S_{6\text{-}140}$ (**17**), was isolated by Ni-affinity chromatography and cleaved with Factor Xa to yield $\alpha S_{6\text{-}140}$ (**18**). We incubated **18** with AaT, Met*RS, tRNA, Hcm, and ATP. Transfer of Hcm was complete after 1 h, within the limits of quantitation by MALDI MS (Figure 3-6). Complete transfer is extremely valuable for large proteins, since separation of an Hcm-modified protein such as **19** from unmodified **18** is not generally feasible. Modified αS **19** was isolated from AaT, Met*RS, and tRNA by boiling and FPLC purification. Then **19** was incubated with thioester Ac-McmMetAspValPhe-SR (**S13**), which corresponds to $\alpha S_{1\text{-}4}$ labeled at the N-terminus with 7-methoxycoumarinylalanine (Mcm). Overnight incubation at 37 °C gave **20a** in high yield from **19**. No unligated starting materials were observed, though some oxidized **20a** was observed. Purified **20a** was quantitatively converted to **21a** by treatment with aqueous MeI, as determined by analyses by John Warner using Ellman's reagent and MALDI MS (Figure 3-7). In general, we expect that the NCL reaction will not be 100 % efficient, but we do not view

this as a severe limitation, provided that the NCL product can be purified away from the starting materials and methylation is quantitative and selective.

A similar ligation was carried out with Ac-MetAspValPhe-SR (**11**), yielding *N*-Ac-αS after methylation (**21b**) (Figure 3-8). The ligation product **21b** was subjected to trypsin digest and MALDI MS analysis to confirm that methylation occurred with high yield and few side reactions. Observation of the Ac-MetAspValPheMetLys fragment (Ac-αS_{1-6}) with only trace Ac-MetAspValPheHcsLys contamination confirmed that methylation proceeded in > 95 % yield within the limitations of quantitation by MS (Figure 3-6). Furthermore, MALDI MS analysis of the rest of the tryptic fragments demonstrated that there were no substantial alkylation side reactions, even on His_{50} in the αS_{46-58} fragment (See Materials and Methods).

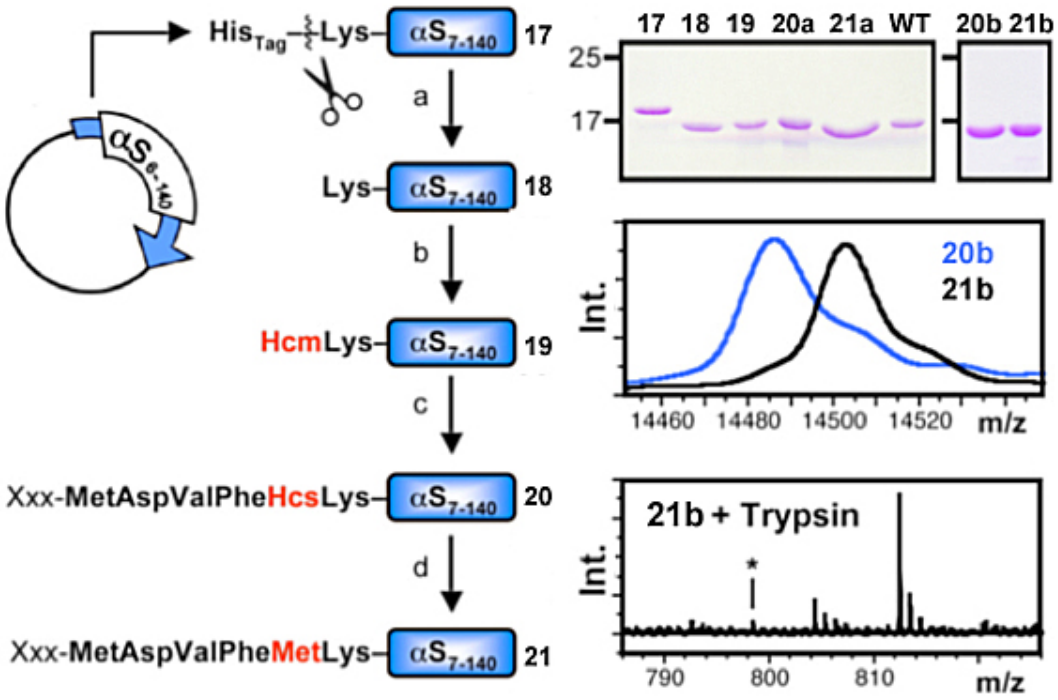


Figure 3-6. α-Synuclein Model Ligation
Functionalization of αS_{6-140} by a) cleavage of His tag at Factor Xa site, b) attachment of Hcm by AaT-catalyzed modification, c) ligation of an N-terminal thioester peptide (**11** or **S10**), and d) conversion of Hcs to Met by methylation. **20a**, **21a** Xxx = AcMcm; **20b**, **21b** Xxx = Ac. Top Right: Gel image showing product of each step, WT indicates full-length αS. Middle Right: MALDI MS analysis of full length Ac-αS before (**12b**) and after (**13b**) methylation. Bottom Right: Trypsinized fragment corresponding to Ac-αS_{1-6} (812.4 m/z) confirms successful methylation. Asterisk indicates the expected mass of unmethylated Ac-$\alpha S_{1-6}Hcs_5$.

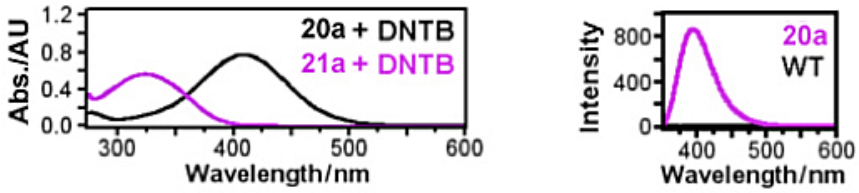


Figure 3-7. Edman Degradation Analysis
Left: UV/Vis analysis of Hcs methylation using Ellman's reagent. Right: Fluorescence emission spectra of Ac-Mcm-αS_{6-140} (**20a**) and WT αS.

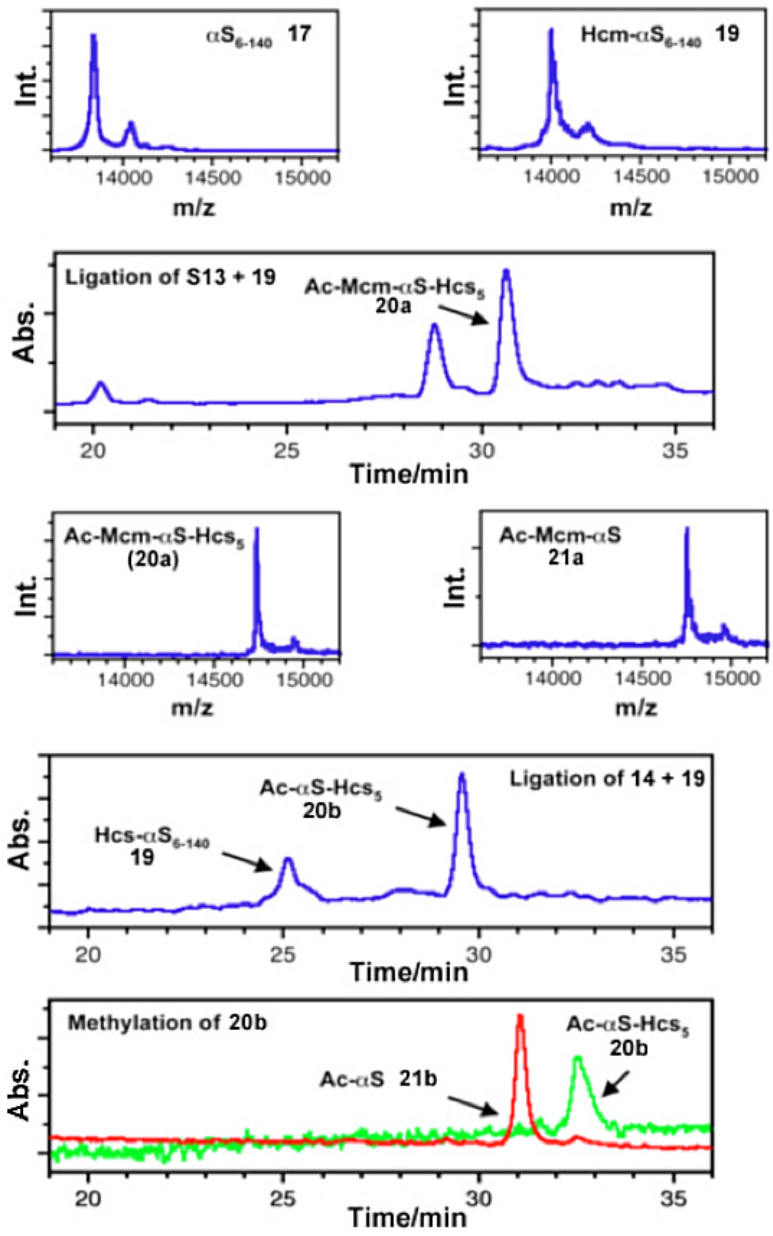
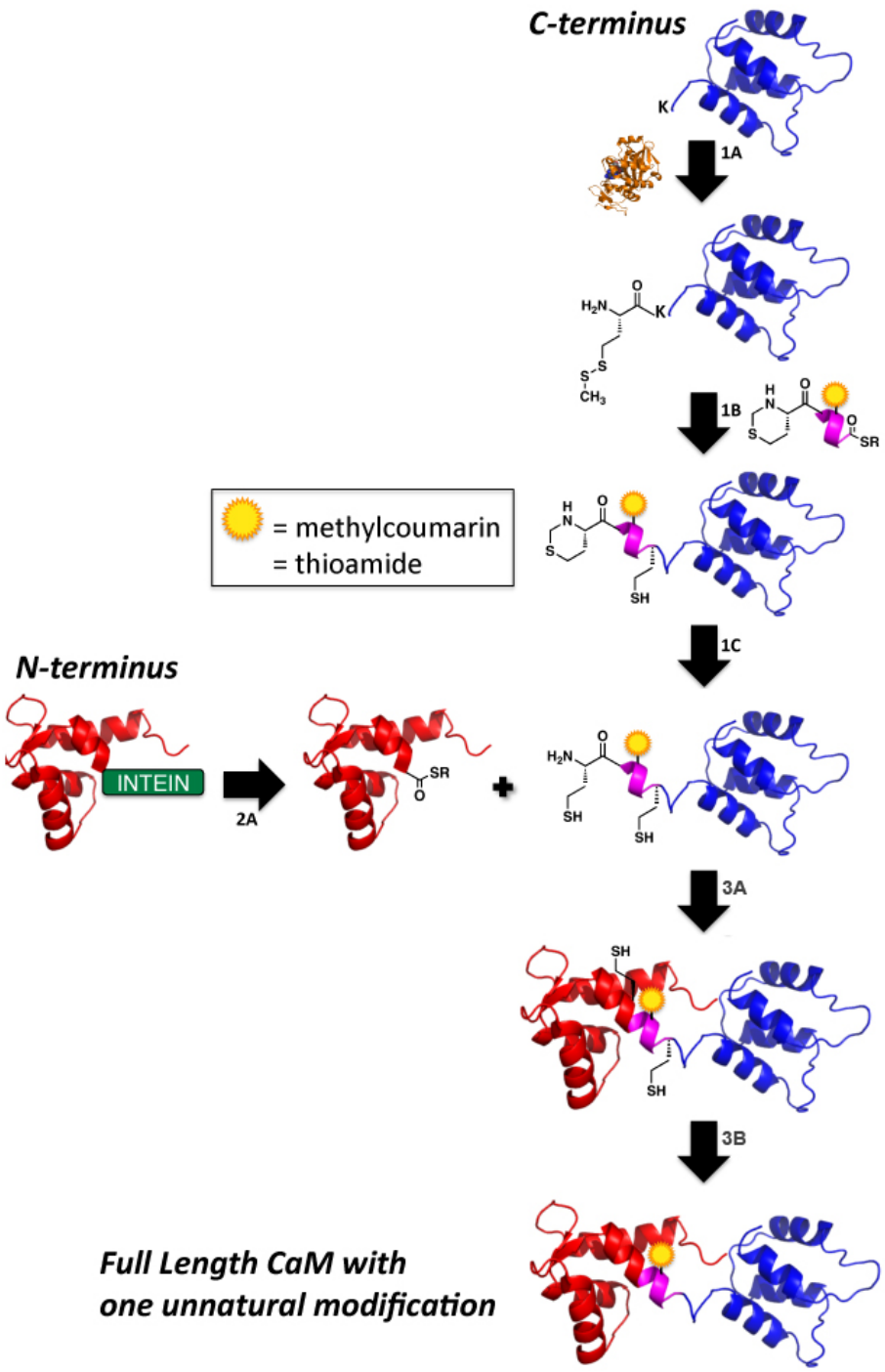


Figure 3-8. HPLC and MALDI MS Analysis of Ligation and Methylation

3.2.4 – Three-Part Native Chemical Ligation with Hcs Masking

With αS, we have shown that we can semi-synthetically create a protein with a new type of ligation site, methionine. To further demonstrate the power of our semi-synthetic techniques, we will perform a three-part native chemical ligation. Our goal is to

introduce one non-native protein element while maintaining native amide bonds at the connection points and retaining the correct primary sequence in the final protein product. As shown in Scheme 3-3, we will recombinantly express the N- and C- terminal domains of calmodulin and only chemically synthesize a small four amino acid middle segment, which contains the non-native element, either a fluorescent label, 7-methoxycoumarinyl alanine, or a thioamide analog of alanine, which can be used as a quencher of Tyr fluorescence in calmodulin. Currently we have reached step 3A, a three-part native chemical ligation with homocysteine at the ligation sites, and have obtained MALDI confirmation of all steps prior.



Scheme 3-3. Three-Part Native Chemical Ligation of Calmodulin
Reaction steps: (1A) AaT N-terminal ligation of Hcm, (1B) Native chemical ligation of middle peptide to C-terminus CaM domain, (1C) Deprotection of thiazolidine to expose N-terminal homocysteine, (2) Cleavage of intein from N-terminus CaM domain to form a thioester, (3A) Native chemical ligation of N-terminus and C-terminus CaM domains, (3B) Methylation of homocysteine residues using methyl iodide.

3.2.5 – Method Applications

Application of our system to targets other than αS may require mild denaturation of the protein in order to access the N-terminal amino acids. Therefore, we have tested AaT activity in the presence of low concentrations of denaturants and detergents. We find that AaT activity has a Gdn•HCl IC_{50} of 0.4 M and that full activity could be maintained in 5 % v/v Triton X-100. While the N-termini of some proteins may be buried in important structural interactions, these concentrations should allow one to partially unfold many proteins to access the terminus for Hcm transfer and ligation.

While our experiments begin to remove the sequence limitations for EPL, they certainly do not eliminate them entirely. Essentially, one can now use XxxMetArg or XxxMetLys as a disconnection point in a retrosynthetic analysis of a potential protein target. The utility of our approach depends on how frequently this motif occurs. To address this, Jerri Wang analyzed the protein sequences in the PDB and found that out of the 60,325 proteins surveyed, 5,405 possible N-terminal ligation sites (< 40 amino acids from the N-terminus) were identified in 5,156 unique proteins. These numbers increased to 31,259 ligation sites and 20,701 proteins when we considered proteins with the ligation motif in the center of the protein where one might use multiple ligations to synthesize the full-length protein. We chose to restrict our search to the PDB since these sequences represent proteins with some degree of prior characterization typical of a protein target chosen for synthesis *via* NCL.

3.2.6 – Future Directions

While it is clear that many proteins could be potential synthesis targets using AaT, we wish to eventually make our methods as broadly useful as possible. Therefore, it is worth noting that there exist other transferases with different N-terminal specificity, such as the Bpt transferase from *V. vulnificus*, which transfers Leu to an N-terminal Asp or Glu.[50] Our preliminary experiments indicate that Bpt can transfer Leu from small molecule donors *in vitro*, but our current yields are too low (< 10 % after 4 h) to be useful. *In situ* aminoacylations using Hcp, Hcb, or Hcm with Bpt and either Met*RS or *E. coli* LeuRS have not produced significant transfer yields. Efforts are underway to improve transfer of Hcs analogs. We are also working to mutate AaT to alter the substrate specificity for both Xaa and the N-terminal residue. (See Chapter 4.)

Prior to our work, the only way to avoid Cys at a ligation site in the expressed protein fragment during EPL was to desulfurize to form Ala or mask Cys by forming a non-native amino acid. Here, we have shown that a protein N-terminus can be functionalized with Hcs under mild conditions. After ligation, Hcs can be efficiently and selectively converted to Met, erasing the ligation site. Our work begins to remove the N-terminal Cys requirement for ligation by allowing one to use the MetArg or MetLys motif as a point of disconnection in protein semi-synthesis.

Section 3.3 – Materials and Methods

3.3.1 – Materials

Homocystine ($(Hcs)_2$) was purchased from Toronto research chemical (Toronto, Canada). Benzotriazol-1-yl-oxy-tris-pyrrolidinophosphonium hexafluorophosphate (PyBOP), Fmoc-Met-OH, Fmoc-Val-OH, Fmoc-Phe-OH, Fmoc-Asp(OtBu)-OH, Fmoc-β-(7-methoxy-coumarin-4-yl)-Ala-OH, and 2-chlorotrityl chloride resin were purchased from Bachem (Torrance, CA) or EMD Chemicals (Philadelphia, PA). Piperidine and 2-(1Hbenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from American Bioanalytical (Natick, MA). Sigmacote, *N,N*-diisopropylethylamine (DIPEA), thiophenol, methyl mercaptopropionate, sodium thiomethoxide, 2-propanthiol, 2-methyl-2-propane thiol, mercaptophenylacetic acid (MPAA), trifluoroacetic acid (TFA), tris(2-carboxyethyl) phosphine hydrochloride (TCEP), and 5,5'-dithiobis-(2-nitrobenzoic acid (DTNB) were purchased from Sigma-Aldrich (St. Louis, MO). Iodine was purchased from JT Baker (Phillipsburg, NJ). All deuterated solvents were purchased from Cambridge Isotopes Laboratories, Inc. (Andover, MA). Ni-NTA resin was from Qiagen (Valencia, CA). *E. coli* BL21(DE3) cells were purchased from Stratagene (La Jolla, CA). Sequencing-grade trypsin was purchased from Promega (Madison, WI). Restriction-grade Factor Xa protease was purchased from Novagen (San Diego, CA). QuikChange® site-directed mutagenesis kits were purchased from Stratagene. DNA oligomers were purchased from Integrated DNA Technologies, Inc (Coralville, IA). Protease inhibitor cocktail was purchased from Sigma-Aldrich. All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). Milli-Q filtered (18 MΩ) water was used for all solutions (Millipore, Billerica,

MA). Matrix-assisted laser desorption/ionization (MALDI) mass spectra were collected with a Bruker Ultraflex III MALDI-TOFTOF mass spectrometer (Billerica, MA). Electrospray ionization (ESI) mass spectra were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA). UV/vis absorbance spectra were obtained with a Hewlett-Packard 8452A diode array spectrophotometer (Agilent Technologies, Santa Clara, CA). Intein plasmid pTXB1 and all restriction enzymes were purchased from New England Biolabs (Ipswich, MA).

3.3.2 – Adenosyl Donor Synthesis

The synthesis of (*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-3-phenylpropanoate (Phe-A, Figure 3-9) has been reported previously.[66] The other Xaa-A donors were synthesized in a similar fashion (Scheme 3-4).

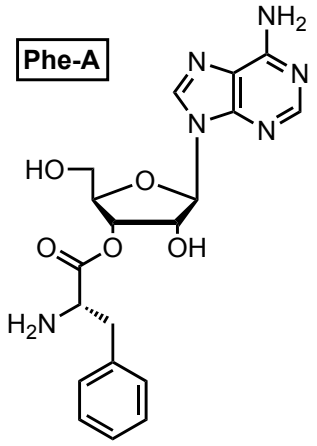
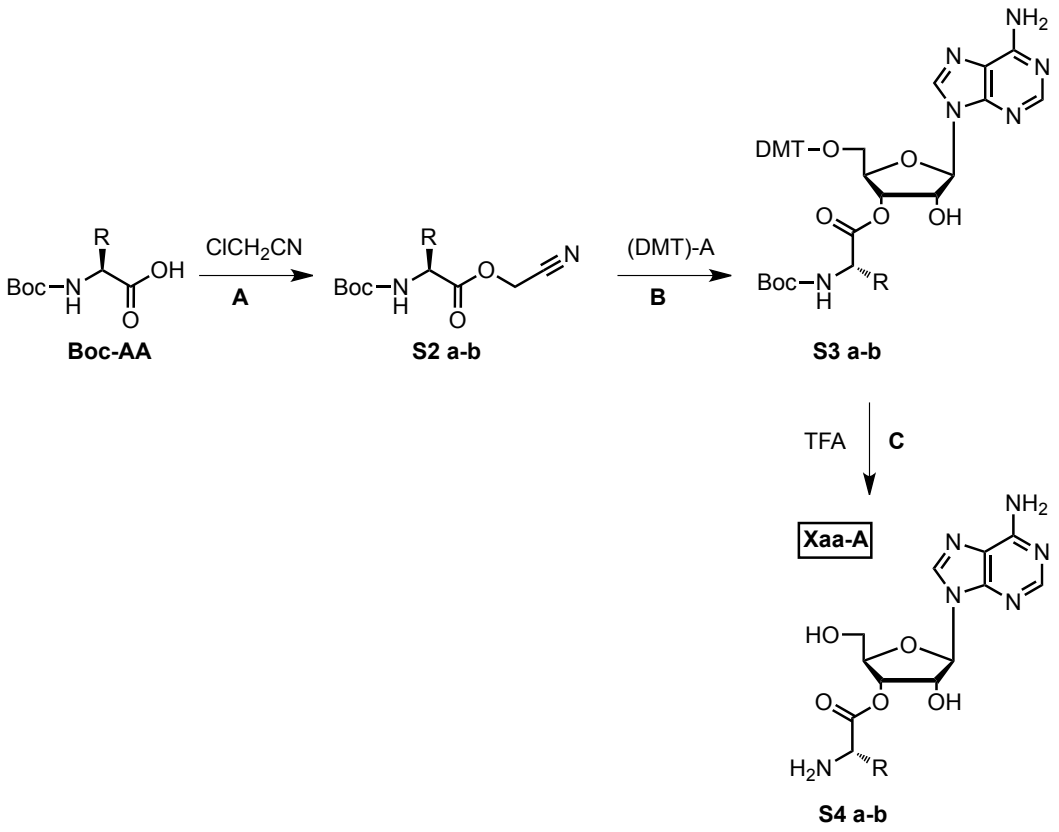


Figure 3-9. Structure of Phenylalanyl Adenosine (Phe-A)



Scheme 3-4. Synthesis of Met-A and Cbz-A Donor
S2-S4a – R = methionine. **S2-S4b** – R = Cys(Bz). Reagents and conditions: a) $ClCH_2CN$, DIPEA, THF; b) (DMT)-A, THF, RT; c) TFA, TIPSH, THF. Cbz-A was synthesized in the same fashion but starting with Boc-Cys(Bz)-OH

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-4-(methylthio)butanoate (Boc-Met-OCH_2CN, S2a).

Chloroacetonitrile (5 mL) and DIPEA (471 mg, 0.64 mL, 3.6 mmol) were added to Boc-Met-OH (826 mg, 3.3 mmol) and stirred for 14 h. The solvent was removed under reduced pressure and SiO_2 flash chromatography (gradient from 10 - 50 % ethyl acetate in hexanes) afforded 799 mg of a pale yellow oil in 83 % yield. R_f 0.45 in 30 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ = 5.19 (d, J = 8.0 Hz, 1H), 4.83 – 4.69 (m, 2H), 4.47 – 4.46 (m, 1H), 2.59 – 2.49 (m, 2H), 2.14 – 1.96 (m, 2H), 2.08 (s,

3H), 1.42 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 171.3, 155.4, 114.1, 80.6, 52.5, 49.1, 31.3, 30.0, 28.4, 15.6,14.3; HRMS (ESI) m/z calculated for $C_{12}H_{20}N_2NaO_4S$ $[M + Na]^+$ 311.104, found 311.104.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-4-(methylthio) butanoate (Boc-Met-(DMT)-A, S3a).

Tetrahydrofuran (5 mL) was added to Boc-Met-OCH_2CN **S2** (219.4 mg, 0.702 mmol), (DMT)-A (179.2 mg, 0.1792 mmol), and NBu_4Ac (7.4 mg) and stirred for 24 h. The solvent was removed under reduced pressure and chromatography (100 % Hexanes, 50 – 100 % ethyl acetate in hexanes, 5 - 10 % methanol in ethyl acetate), afforded 92.9 mg of an off-white solid in 64.7 % yield. R_f 0.5 in 5 % methanol in ethyl acetate; ^{1}H and ^{13}C NMR for **S3a** shown in Figure 3-10; LRMS (ESI) m/z calculated for $C_{41}H_{49}N_6O_9S$ $(M + H)^+$ 801.3, found 801.6.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-4-(methylthio)butanoate (Met-A, S4a).

Trifluoroacetic acid (1 mL tetrahydrofuran (1 mL) and TIPSH (73.5 mg, 0.095 mL, 0.46 mmol) were added to **S3a** (92.9 mg, 0.116 mmol). Upon addition of TFA, the reaction turned a bright orange color and at the completion of the reaction, the color had become a pale yellow - orange. The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a final

1 mL water back-extraction against the dichloromethane layer. The water-soluble layer containing **S4a** was then HPLC-purified on a C18-prep column using Gradient 1 at 11 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{15}H_{23}N_6O_5S$ $(M + H)^+$ 399.1; retention time 21.0 min, found 399.3 retention time 22.44 min, found 399.3.

(*R*)-cyanomethyl 3-(benzylthio)-2-((*tert*-butoxycarbonyl)amino)propanoate (Boc-Cbz-OCH_2CN, S2b).

Boc-Cys(benzyl)-OH (500.0 mg, 1.681 mmol) was dissolved in 5 mL chloroacetonitrile (solvent) and DIPEA (0.322 mL, 1.849 mmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure, then purified on silica gel (30 % EtOAc in hexanes) to give 588.7 mg (~ 100 %) of the titled compound **S2b**. R_f 0.4 in 30 % EtOAc in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.30 – 7.20 (m, 5H), 4.68 (s, 2H), 4.51 – 4.5 (d, 1H), 3.67 (s, 2H), 2.85 – 2.75 (m, 2H). 1.40 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 170.1, 155.2, 137.5, 129.1, 128.9, 127.7, 113.9, 80.9, 53.2, 49.3, 36.9, 33.3, 28.5. LRMS (ESI) m/z calculated for $C_{17}H_{22}N_2O_4SNa$ $[M + Na]^+$ 373.1, found 373.2.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 3-(benzylthio)-2-((*tert*-butoxycarbonyl)amino)propanoate (Boc-Cbz-(DMT)-A, S3b).

Tetrahydrofuran (5 mL) was added to Boc-Cys(Bz)-OCH_2CN **S2b** (279.4 mg,

0.797 mmol), (DMT)-A (113 mg, 0.199 mmol), and NBu_4Ac (6.3 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate, eluted in 10 % methanol in chloroform), afforded 42.3 mg of a white solid in 24.6 % yield. ^1H and ^{13}C NMR for (Figure 3-11) shown below; LRMS (ESI) m/z calculated for $C_{46}H_{51}N_6O_9S$ $(M + H)^+$ 863.3, found 863.5.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(benzylthio)propanoate (Cbz-A, S4b).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (7.8 mg, 0.01 mL, 0.05 mmol) were added to **S3b** (28.5 mg, 0.0331 mmol) following the deprotection procedure given for **S4a**. The water-soluble layer containing **S3b** was then HPLC-purified on a C18-prep column using Gradient 1 at 15 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{20}H_{25}N_6O_5S$ $(M + H)^+$ 461.2; retention time 17.5 min, found 460.5.

(*R*)-2-((*tert*-butoxycarbonyl)amino)-3-(methyldisulfanyl)propanoic acid (Boc-Csm-OH, 7a).

Boc-L-Cys-OH (110 mg, 0.50 mmol) was dissolved in 20 mL of 1:1 H_2O/THF. Sodium thiomethoxide (350 mg, 5.0 mmol) was added to the solution after cooling to 0 °C in an ice bath. Then, a solution of iodine in 95 % EtOH was added dropwise until the

color of the reaction changed from colorless to brown. The mixture was stirred for 5.5 h with warming to room temperature. After removing tetrahydrofuran under reduced pressure, the reaction mixture was acidified by addition of 1 N HCl aq. until the pH was 2 - 3. The solution was extracted with EtOAc. The organic phase was washed with saturated NaS_2O_4 aq. and brine, and dried over $MgSO_4$. The solvent was removed under reduced pressure, and then the reaction was purified on silica gel (25 % EtOAc in hexanes containing 1 % AcOH) to give 90.8 mg (68 %) of the titled compound **7a**. R_f 0.37 in 33 % ethyl acetate in hexanes containing 1 % AcOH; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.34 (s, 1H), 4.63 (s, 1H), 3.26 - 3.16 (m, 2H), 2.43 (m, 3H), 1.47 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 175.2, 155.7, 80.9, 53.2, 39.9, 28.5, 23.3. NMRs, ^{1}H and ^{13}C, can be found in Figure 3-12. HRMS (ESI) m/z calculated for $C_9H_{17}NNaO_4S_2$ $[M + Na]^+$ 290.049, found 290.050.

(*R*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-3-(methyldisulfanyl)propanoate (Boc-Csm-OCH$_2$CN, 8a).

Compound **7a** (106 mg, 0.397 mmol) was dissolved in 3 mL tetrahydrofuran. 50 equiv. chloroacetonitrile (1.26 mL, 19.85 mmol) and DIPEA (67 μL, 796 μmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure, then purified on silica gel (25 % EtOAc in hexanes) to give 47 mg of compound **8a** (44 %). R_f 0.33 in 25 % EtOAc in hexanes; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.34 (s, 1H), 4.87 – 4.74 (dd, J = 33.1 Hz, 2H), 4.71 (s, 1H), 3.18 - 3.11 (m, 2H), 2.43 (m, 3H), 1.46 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 170.0, 155.2,

113.9, 81.0, 52.8, 49.3, 39.2, 28.5, 23.1; HRMS (ESI) m/z calculated for $C_{11}H_{18}N_2NaO_4S_2$ $[M + Na]^+$ 392.061, found 392.061.

(*R*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(methyldisulfanyl) propanoate (Boc-Csm-(DMT)-A, 9a).

Tetrahydrofuran (5 mL) was added to Boc-Csm-OCH_2CN **8a** (123.6 mg, 0.403 mmol), (DMT)-A (115.7 mg, 0.2031 mmol), and NBu_4Ac (7.4 mg,) and stirred for 48 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate and eluted from silica with 10 % methanol in chloroform), afforded 80.8 mg of a white solid in 48.6 % yield. R_f 0.7 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR shown below in Figure 3-13; LRMS (ESI) m/z calculated for $C_{40}H_{47}N_6O_9S_2$ $(M + H)^+$ 819.3, found 819.4.

(*R*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(methyldisulfanyl)propanoate (Csm-A, 10a).

Trifluoroacetic acid (1mL), tetrahydrofuran (1 mL) and TIPSH (62.6 mg, 0.081 mL, 0.40 mmol) were added to **9a** (80.8 mg, 0.0987 mmol) following the deprotection procedure given for **S4a**. The water-soluble layer containing **10a** was then HPLC-purified on a C18-prep column using Gradient 1 at 11 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{14}H_{21}N_6O_5S_2$ $(M + H)^+$ 417.1; retention time

21.3 min, found 416.8; retention time 23.4 min, found 416.8.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl) (phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(isopropyldisulfanyl)propanoate (Boc-Csp-(DMT)-A, 9b).

Tetrahydrofuran (5 mL) was added to Boc-Csp-OCH_2CN **8b** (332.5 mg, 0.994 mmol), (DMT)-A (141.5 mg, 0.249 mmol), and NBu_4Ac (12.5 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and flash chromatography over silica gel (Gradient 20 % -100 % ethyl acetate in hexanes, 5 % - 10 % methanol in ethyl acetate) afforded 159.3 mg of a white solid in 75.7 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for (Figure 3-14) shown below; LRMS (ESI) m/z calculated for $C_{42}H_{51}N_6O_9S_2$ $(M + Na)^+$ 869.3, found 869.6.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(isopropyldisulfanyl)propanoate (Csp-A, 10b).

Trifluoroacetic acid (1mL), tetrahydrofuran (1 mL) and TIPSH (11.9 mg, 0.154 mL, 0.752 mmol) were added to **9b** (159.3 mg, 0.188 mmol) following the deprotection procedure given for **S4a**. The water-soluble layer containing **10b** was then HPLC-purified on a C18-prep column using Gradient 1 at 11 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{16}H_{25}N_6O_5S_2$ $(M + H)^+$ 445.1; retention time 26.1 min, found 445.2; retention time 28.1 min, found 445.2.

(*R*)-2-((*tert*-butoxycarbonyl)amino)-3-(*tert*-butyldisulfanyl)propanoic acid (Boc-Csb-OH, 7c).

Cys(*S*-*t*Bu)-OH (250 mg, 1.2 mmol) was dissolved in 12 mL of 1:1 H_2O/dioxane. Sodium hydroxide (71.6 mg, 1.8 mmol) and di-*tert*-butyl carbonyl anhydride (391 mg, 1.8 mmol) were added to the solution and the mixture was stirred overnight at room temperature. 20 mL of MilliQ water was added to the reaction, and then the reaction mixture was extracted twice with ethyl acetate. The organic layer was back extracted twice with saturated sodium bicarbonate solution. The combined aqueous layers were acidified to a pH of 1 using 6 M HCl aq. The solution was extracted with EtOAc three times. The extract was washed with brine, and dried over $MgSO_4$. After filtration, solvent was removed to give 263.4 mg (70 %) of **7c**, a white solid. Further purification was not performed. R_f 0.5 in 60 % ethyl acetate in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): 4.59 - 4.47 (m, 1H), 3.23 – 3.13 (m, 2H), 1.45 (s, 9H), 1.32 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 175.5, 155.6, 80.8, 53.4, 48.4, 42.3, 30.0, 28.5; LRMS (ESI) m/z calculated for $C_{12}H_{23}NO_4S_2Na$ $(M + Na)^+$ 332.1, found 332.2.

(*R*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-3-(*tert*-butyldisulfanyl)propanoate (Boc-Csb-OCH$_2$CN, 8c).

Compound **7c** (206.7 mg, 0.668 mmol) was dissolved in 5 mL chloroacetonitrile (solvent) and DIPEA (0.128 mL, 0.735 mmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure, then purified on silica gel (30 % EtOAc in hexanes) to give 140.6 mg (60 %) of the titled

compound **8c**. R_f 0.6 in 30 % EtOAc in hexanes; ^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.80 (dd, *J* = 12.5 Hz, 2H), 2.66 - 2.64 (m, 1H), 3.15 – 3.14 (m, 2H), 1.44 (s, 9H), 1.33 (s, 9H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 169.9, 155.2, 114.0, 80.8, 53.1, 49.3, 48.8, 41.9, 29.9, 28.4; LRMS (ESI) m/z calculated for $C_{14}H_{24}N_2O_4S_2Na$ $[M + Na]^+$ 371.1, found 371.2.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl) methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(*tert*-butyldisulfanyl)propanoate (Boc-Csb-(DMT)-A, 9c).

Tetrahydrofuran (5 mL) was added to Boc-Csb-OCH_2CN **8c** (140.6 mg, 0.404 mmol), (DMT)-A (57.5 mg, 0.101 mmol), and NBu_4Ac (7.7 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate, eluted in 10 % methanol in chloroform), afforded 28.5 mg of a white solid in 30.4 % yield. R_f 0.5 in 5 % methanol in ethyl acetate; ^{1}H and ^{13}C NMR for (Figure 3-15) shown below; LRMS (ESI) m/z calculated for $C_{43}H_{52}N_6O_9S_2Na$ $(M + Na)^+$ 883.3, found 883.6.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(*tert*-butyldisulfanyl)propanoate (Csb-A, 10c).

Trifluoroacetic acid (1mL), tetrahydrofuran (1 mL) and TIPSH (21 mg, 0.027 mL, 0.13 mmol) were added to **9c** (28.5 mg, 0.0331 mmol) following the deprotection

procedure given for **S4a**. The water-soluble layer containing **10c** was then HPLC-purified on a C18-prep column using Gradient 1 with a flow rate of 15 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{17}H_{27}N_6O_5S_2$ $(M + H)^+$ 459.2; retention time 20.0 min, found 459.0; retention time 22.1 min, found 459.0.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*t*-butoxycarbonyl)amino)-3-(phenyldisulfanyl)propanoate (Boc-Csf-(DMT)-A, 9d).

Tetrahydrofuran (5 mL) was added to Boc-Csf-OCH_2CN **8d** (110.0 mg, 0.303 mmol), (DMT)-A (85.0 mg, 0.152 mmol), and NBu_4Ac (15 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate, eluted in 10 % methanol in chloroform), afforded 64.5 mg of a yellow solid in 48.3 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for (Figure 3-16) shown below; LRMS (ESI) m/z calculated for $C_{45}H_{49}N_6O_9S_2$ $(M + H)^+$ 881.3, found 881.4.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(phenyldisulfanyl)propanoate (Csf-A, 10d).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (46.4 mg, 0.06 mL, 0.29 mmol) were added to **9d** (64.5 mg, 0.073 mmol) following the deprotection

procedure given for **S4a**. The water-soluble layer containing **10d** was then HPLC-purified on a C18-prep column using Gradient 1 at 11 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{19}H_{23}N_6O_5S_2$ $(M + H)^+$ 479.1; retention time 29.1 min, found 479.0; retention time 30.6 min, found 479.0.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-((2,2,2-trifluoroethyl)disulfanyl)propanoate (Boc-CsTf-(DMT)-A, 9e).

Tetrahydrofuran (5 mL) was added to Boc-Csf-OCH_2CN **8e** (36.9 mg, 0.10 mmol), (DMT)-A (40.7 mg, 0.07 mmol), and NBu_4Ac (2.7 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate, eluted in 10 % methanol in chloroform), afforded 18.8 mg of a white solid in 29.7 % yield. R_f 0.8 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for (Figure 3-17) shown below; LRMS (ESI) m/z calculated for $C_{41}H_{46}F_3N_6O_9S_2$ $(M + H)^+$ 887.3, found 887.5.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-((2,2,2-trifluoroethyl)disulfanyl)propanoate (CsTf-A, 10d)

Trifluoroacetic acid (1mL), tetrahydrofuran (1 mL) and TIPSH (13.4 mg, 0.02 mL, 0.085 mmol) were added to **9e** (18.8 mg, 0.021 mmol) following the deprotection

procedure given for **S4a**. The water-soluble layer containing **10d** was then HPLC-purified on a C18-prep column using Gradient 1 at 11 mL/min (Table 3-1). HPLC/MALDI analysis m/z calculated $C_{15}H_{20}F_3N_6O_5S_2$ $(M + H)^+$ 485.1; retention time 25.4 min, found 485.0; retention time 27.5 min, found 485.0.

(2*S*,2'*S*)-4,4'-disulfanediylbis(2-((*tert*-butoxycarbonyl)amino)butanoic acid ((Boc-Hcs)$_2$, 1).

L-Homocystine ($(Hcs)_2$, 402.5 mg, 1.5 mmol) was dissolved in 15 mL of 1:1 H_2O/dioxane. 600 mL of 10 N Sodium hydroxide aq. (6 mmol) and di-*tert*-butyl carbonyl anhydride (618 mg, 3.75 mmol) were added to the solution after cooling to 0 ºC in the ice bath. The mixture was stirred overnight with warming to room temperature. The reaction mixture was acidified by 1 N HCl aq. until reached 2 - 3. The solution was extracted with EtOAc. The extract was washed with brine, and dried over $MgSO_4$. After filtration, solvent was removed to give 572.8 mg (81 %) of **1**. Further purification was not performed.

(*S*)-2-((*tert*-butoxycarbonyl)amino)-4-(methyldisulfanyl)butanoic acid (Boc-Hcm-OH, 2a).

Boc-L-Homocystine (**1**), 188 mg, 0.4 mmol) was dissolved in 16 mL of 1:1 H_2O/THF. Sodium thiomethoxide (560 mg, 8.0 mmol) was added to the solution after cooling to 0 °C in an ice bath. Then, a saturated solution of iodine in 95 % EtOH was added dropwise until the color of the reaction changed from colorless to brown. The

mixture was stirred for 1.5 h with warming to room temperature. After removing tetrahydrofuran under reduced pressure, the reaction mixture was acidified by addition of 1 N HCl aq. until the pH reached 2 - 3. The solution was extracted with EtOAc. The extract was washed with saturated NaS_2O_4 aq. and brine, dried over $MgSO_4$, and purified on silica gel (25 % EtOAc in hexanes containing 1 % AcOH) to give 179.3 mg (80 %) of **2a**. R_f 0.4 in 33 % ethyl acetate in hexanes containing 1 % AcOH; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.34 (s, 1H), 4.63 (s, 1H), 2.80 - 2.76 (m, 2H), 2.42 (s, 2H), 2.35 (m, 1H), 2.09 (s, 1H), 1.46 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 176.5, 155.6, 80.6, 52.5, 33.8, 32.8, 28.3, 23.3 (Note: A second set of concentration-dependent peaks was also observed, these peaks can be seen in the included ^{13}C spectrum. These could be reduced but not eliminated by diluting the sample, no evidence of dimerization is seen in the more dilute ^{1}H spectrum sample, Figure 3-18); HRMS (ESI) m/z calculated for $C_{10}H_{19}NNaO_4S_2$ $[M + Na]^+$ 304.065, found 304.066.

(*S*)-2-amino-4-(methyldisulfanyl)butanoic acid (Hcm-OH, free amino acid)

Boc-Hcm-OH (**2a**), 150 mg, 0.53 mmol) was dissolved in 10 mL of 95:2.5:2.5 TFA/H_2O/TIPS. Reaction was stirred for 2 hours. Solvent was removed under reduced pressure. Crude Hcm was extracted using excess cold diethylether. Precipitate was collected *via* centrifugation and pellet was resuspended in water. Crude Hcm was purified by HPLC (Gradient 2). Post purification, sample concentration was determined by UV-Vis at 280 nm using the extinction coefficient of disulfide (ε_{280} = 125 M^{-1} cm^{-1}). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 4.12 (t, *J* = 6.5 Hz, 1H), 2.95 (t, *J* = 7.5 Hz, 2H), 2.52

(s, 1H), 2.50 – 2.39 (m, 2H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 173.0, 52.7, 32.0, 29.6, 22.1; HRMS (ESI) m/z calculated for $C_5H_{12}NO_2S_2$ $[M + H]^+$ 182.030, found 182.031. NMRs can be found in Figure 3-19.

(*S*)-2-((*tert*-butoxycarbonyl)amino)-4-(isopropyldisulfanyl)butanoic acid (Boc-Hcp-OH, 2b).

1 (234 mg, 0.5 mmol) was dissolved in 19 mL of 50 % aqueous tetrahydrofuran. 2-propanethiol (929 μL, 10.0 mmol) and 10 N sodium hydroxide aq. (1.1 mL, 11 mmol) were added to the solution at 0 °C. Then, a solution of iodine in 95 % EtOH was added dropwise until the color of the reaction system changed from colorless to brown. The mixture was stirred for over-night with gradual warming room temperature. After removing THF under reduced pressure, reaction mixture was neutralized by 1 N HCl aq. until pH 2 - 3. The solution was extracted with EtOAc. The extract was washed with saturated NaS_2O_4 aq. and brine, and dried over $MgSO_4$. Concentration under reduced pressure followed by flash chromatography over silica gel with EtOAc/Hexane/AcOH (75:25:1) gave **2b** (170 mg, 55 % yield). R_f 0.47 in 33 % ethyl acetate in hexanes containing 1 % AcOH; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.05 (s, 1H), 4.40 (s, 1H), 3.07 – 2.95 (m, 1H), 2.75 (t, J = 7.6 Hz, 2H), 2.30 (s, 1H), 2.08 (s, 1H), 1.46 (s, 9H), 1.32 - 1.30 (dd, J = 6.8 Hz, 6H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 176.6, 155.8, 80.7, 52.8, 41.3, 35.6, 32.5, 28.5, 22.8 (Note: A second set of concentration-dependent peaks was also observed, these peaks can be seen in the included ^{13}C spectrum. These could be reduced, but not eliminated by diluting the sample, no evidence of dimerization is seen in the more

dilute ^{1}H spectrum sample, Figure 3-21); HRMS (ESI) m/z calculated for $C_{12}H_{23}NNaO_4S_2$ $[M + Na]^+$ 332.096, found 332.097.

(*S*)-2-((*tert*-butoxycarbonyl)amino)-4-(*tert*-butyldisulfanyl)butanoic acid (Boc-Hcb-OH, 2c).

1 (56.4 mg, 0.12 mmol) was dissolved in H_2O/THF (50 % v/v, 19 mL). To the solution was added 2-methyl-2-propanethiol (270 μL, 2.4 mmol) and 5 N sodium hydroxide aq. (528 μL, 2.64 mmol) after cooling to 0 °C in an ice bath. Then, a solution of iodine in 95 % EtOH was added dropwise until the color of the reaction changed from colorless to brown. The mixture was stirred overnight with warming to room temperature. After removing THF under reduced pressure, the reaction mixture was acidified by addition of 1 N HCl aq. until the pH reached 2 - 3. The solution was extracted with EtOAc. The extract was washed with saturated NaS_2O_4 aq. and brine, and dried over $MgSO_4$. Concentration under reduced pressure followed by flash chromatography over silica gel with EtOAc/Hexane/AcOH (75:25:1) gave **2c** (78.4 mg, quant. yield). R_f 0.5 in 33 % ethyl acetate in hexanes containing 1 % AcOH; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.05 (s, 1H), 4.39 (s, 1H), 2.78 - 2.74 (m, 2H), 2.30 (s, 1H), 2.10 (s, 1H), 1.46 (s, 9H), 1.33 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 176.2, 155.7, 114.0, 80.4, 60.7, 52.8, 48.0, 33.6, 31.9, 28.5, 23.3 (Note: A second set of concentration-dependent peaks was also observed; these peaks can be seen in the included ^{13}C spectrum. These could be reduced, but not eliminated by diluting the sample, no evidence of dimerization is seen in the more dilute ^{1}H spectrum sample, Figure 3-23);

HRMS (ESI) m/z calculated for $C_{13}H_{24}NO_4S_2$ $[M + H]^+$ 322.115, found 322.115.

(*S*)-2-((*tert*-butoxycarbonyl)amino)-4-(phenyldisulfanyl)butanoic acid (Boc-Hcf-OH, 2d).

1 (234 mg, 0.5 mmol) was dissolved in H_2O/THF (50 % v/v, 19 mL). To the solution was added thiophenol (600 μL, 5.9 mmol) and 10 N sodium hydroxide aq. (700 μL, 7 mmol) after cooling to 0 °C in an ice bath. Then, a solution of iodine in 95 % EtOH was added dropwise until the color of the reaction changed from colorless to brown. The mixture was stirred overnight with warming to room temperature. After removing THF under reduced pressure, the reaction mixture was acidified by addition of 1 N HCl aq. until the pH reached 2 - 3. The solution was extracted with EtOAc. The extract was washed with saturated NaS_2O_4 aq. and brine, and dried over $MgSO_4$. Concentration under reduced pressure followed by flash chromatography over silica gel with EtOAc/Hexane/AcOH (80:20:1) gave **2d** (166 mg, 44 %). R_f 0.47 in 33 % ethyl acetate in hexanes containing 1 % AcOH ^{1}H NMR (360 MHz, $CDCl_3$): δ = 7.52-7.50 (m, 2H), 7.32-7.28 (m, 2H), 7.23-7.21 (m, 1H), 4.98 (s, 1H), 4.36 (s, 1H), 2.80 - 2.76 (m, 2H), 2.35-2.29 (m, 1H), 2.08-2.00 (m, 1H), 1.42 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 176.5, 155.8, 137.2, 129.3,128.2, 127.3, 80.8, 52.7, 34.6, 32.1, 28.5 (Note: A second set of concentration-dependent peaks was also observed, these peaks can be seen in the included ^{13}C spectrum. These could be reduced, but not eliminated by diluting the sample, no evidence of dimerization is seen in the more dilute ^{1}H spectrum sample, Figure 3-25); HRMS (ESI) m/z calculated for $C_{15}H_{21}NNaO_4S_2$ $[M + Na]^+$ 366.080, found

366.081.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-4-(methyldisulfanyl)butanoate (Boc-Hcm-OCH$_2$CN, 3a).

Compound **2a** (367 mg, 1.3 mmol) was dissolved in 4 mL tetrahydrofuran. 50 equiv. chloroacetonitrile (4.1 mL, 65.2 mmol) and DIPEA (453 mL, 2.6 mmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure, then purified on silica gel (25 % EtOAc in hexanes) to give 307.5 mg (74 %) of the titled compound **3a**. R_f 0.33 in 25 % EtOAc in hexanes; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.02 (s, 1H), 4.73 (dd, J = 31.3 Hz, 2H), 4.49 (s, 1H), 2.77 - 2.73 (m, 2H), 2.37 - 2.27 (m, 1H), 2.13 - 2.02 (m, 1H), 1.46 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 171.2, 155.4, 114.03, 80.9, 52.5, 49.2, 33.6, 31.9, 28.5, 23.3; HRMS (ESI) m/z calculated for $C_{12}H_{20}N_2NaO_4S_2$ $[M + Na]^+$ 343.076, found 343.076.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-4-(isopropyldisulfanyl)butanoate (Boc-Hcp-OCH$_2$CN, 3b).

Compound **2b** (100 mg, 324 μmol) was dissolved in 1 mL tetrahydrofuran. 50 equiv. chloroacetonitrile (1.08 mL, 16.2 mmol) and DIPEA (113 uL, 647 μmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure, then purified on silica gel (25 % EtOAc in hexanes) to give 95.1 mg (84 %) of **3b**. R_f 0.43 in 25 % EtOAc in hexanes; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.01 (s, 1H), 4.85 - 4.73 (m, 2H), 4.46 (s, 1H), 3.03 - 2.99 (m, 1H), 2.72 (t,

2H), 2.30 - 2.27 (m, 1H), 2.09 - 2.03 (m, 1H), 1.45 (s, 9H), 1.33 - 1.31 (m, 4H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 171.2, 155.4, 114.0, 80.8, 52.5, 49.2, 41.3, 35.2, 31.9, 28.4, 22.8, 22.8; HRMS (ESI) m/z calculated for $C_{14}H_{24}N_2NaO_4S_2$ $[M+Na]^+$ 371.107, found 371.108.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-4-(*tert*-butyldisulfanyl)butanoate Boc-Hcb-OCH$_2$CN, 3c).

Compound **2c** (100 mg, 309 μmol) was dissolved in 1 mL tetrahydrofuran. 50 equiv. chloroacetonitrile (977 μL, 15.5 mmol) and DIPEA (108 uL, 618 μmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure, then purified on silica gel (25 % EtOAc in hexanes) to give 106.3 mg (95 %) of compound **3c**. R_f 0.47 in 25 % EtOAc in hexanes; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 5.00 (s, 1H), 4.73 (dd, J = 25.2 Hz, 2H), 4.46 (s, 1H), 2.74 (t, 2H), 2.33 - 2.24 (m, 1H), 2.11 - 2.00 (m, 1H), 1.45 (s, 9H), 1.34 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 171.2, 155.4, 114.1, 80.8, 52.6, 49.2, 48.3, 35.9, 31.9, 30.1, 28.5; HRMS (ESI) m/z calculated for $C_{15}H_{26}N_2NaO_4S_2$ $[M+Na]^+$ 385.123, found 385.124.

(*S*)-cyanomethyl 2-((*tert*-butoxycarbonyl)amino)-4-(phenyldisulfanyl)butanoate (Boc-Hcf-OCH$_2$CN, 3d).

Compound **2d** (100 mg, 291 μmol) was dissolved in 1 mL tetrahydrofuran. 50 equiv. chloroacetonitrile (921 μL, 14.6 mmol) and DIPEA (101.5 uL, 583 μmol) were added to the reaction mixture and allowed to stir overnight. The reaction mixture was

concentrated under reduced pressure, then purified on silica gel (25 % EtOAc in hexanes) to give 103.8 mg (93 %) of compound **3d**. R_f 0.37 in 25 % EtOAc in hexanes; ^{1}H NMR (360 MHz, $CDCl_3$): δ = 7.53-7.50 (m, 2H), 7.34-7.30 (m, 2H), 7.25-7.21 (m, 1H), 4.91 (s, 1H), 4.73 (dd, *J* = 25.2 Hz, 2H), 4.41 (s, 1H), 2.79-2.74 (m, 2H), 2.30 - 2.22 (m, 1H), 2.07 - 2.01 (m, 1H), 1.42 (s, 9H); ^{13}C NMR (90 MHz, $CDCl_3$): δ = 171.1, 155.4, 137.0, 129.4, 128.3, 127.5, 114.0, 80.9, 52.4, 49.2, 34.5, 31.5, 28.5; HRMS (ESI) m/z calculated for $C_{17}H_{21}N_2O_4S_2$ $[M + H]^+$ 381.094, found 381.094.

***(S)-(2R,3R,4R,5R)*-2-(6-amino-*9H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((tert-butoxycarbonyl)amino)-4-(methyldisulfanyl) butanoate (Boc-Hcm-(DMT)-A, 4a).**

Tetrahydrofuran (5 mL) was added to Boc-Hcm-OCH_2CN **3a** (80 mg, 0.250 mmol), (DMT)-A (71.3 mg, 0.125 mmol), and NBu_4Ac (5 mg) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate and eluted from silica with 10 % methanol in chloroform), afforded 67.4 mg of a white solid in 64.7 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^{1}H and ^{13}C NMR for **4a** shown below in Figure 3-20; LRMS (ESI) m/z calculated for $C_{41}H_{49}N_6O_9S_2$ $(M + H)^+$ 833.3, found 833.4.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy) methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-4-(isopropyldisulfanyl) butanoate (Boc-Hcp-(DMT)-A, 4b).

Tetrahydrofuran (5 mL) was added to Boc-Hcp-OCH_2CN **3b** (74.5 mg, 0.214 mmol), (DMT)-A (61.2 mg, 0.107 mmol), and NBu_4Ac (4 mg) and stirred for 24 h. The solvent was removed under reduced pressure and chromatography (100 % hexanes, 20 – 100 % ethyl acetate in hexanes, 2.5 - 10 % methanol in ethyl acetate), afforded 38.0 mg of a white solid in 41.1 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for **4b** shown below, Figure 3-22; LRMS (ESI) m/z calculated for $C_{43}H_{53}N_6O_9S_2$ $(M + H)^+$ 861.3, found 861.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-4-(*tert*-butyldisulfanyl)butanoate (Boc-Hcb-(DMT)-A, 4c).

Tetrahydrofuran (5 mL) was added to Boc-Hcb-OCH_2CN **3c** (95.3 mg, 0.263 mmol), (DMT)-A (76.6 mg, 0.1316 mmol), and NBu_4Ac (11.4 mg) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin layer chromatography (5 % methanol in ethyl acetate and eluted from silica with 10 % methanol in dichloromethane), afforded 94.6 mg of a white solid in 39.9 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for **4c** shown below, Figure 3-24; LRMS (ESI) m/z calculated for $C_{44}H_{55}N_6O_9S_2$ $(M + H)^+$ 875.3, found 875.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-4-(phenyldisulfanyl)butanoate (Boc-Hcf-(DMT)-A, 4d).

Tetrahydrofuran (5 mL) was added to Boc-Hcf-OCH_2CN **3d** (103.8 mg, 0.272 mmol), (DMT)-A (78.3 mg, 0.137 mmol), and NBu_4Ac (13.5 mg) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin layer chromatography (5 % methanol in ethyl acetate and eluted from silica with 10 % methanol in dichloromethane), afforded 91.5 mg of a white solid in 74.6 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for **4d** shown below, Figure 3-26; LRMS (ESI) m/z calculated for $C_{46}H_{51}N_6O_9S_2$ $(M + H)^+$ 895.3, found 895.5.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-4-((2,2,2-trifluoroethyl)disulfanyl)butanoate (Boc-HcTf-(DMT)-A, 4e).

Tetrahydrofuran (5 mL) was added to Boc-HcTf-OCH_2CN **3e** (76.1 mg, 0.196 mmol), (DMT)-A (75.9 mg, 0.133 mmol), and NBu_4Ac (4.7 mg) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin layer chromatography (5 % methanol in ethyl acetate and eluted from silica with 10 % methanol in chloroform), afforded 94.9 mg of a white solid in 79.1 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR for **4e** shown below, Figure 3-27; LRMS (ESI) m/z calculated for $C_{42}H_{48}F_3N_6O_9S_2$ $(M + H)^+$ 901.3, found 901.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-4-(methyldisulfanyl)butanoate (Hcm-A, 5a).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (51.3 mg, 0.066 mL, 0.32 mmol) were added to **4a** (67.4 mg, 0.081 mmol) following the deprotection procedure given for **S4a**. HPLC/MALDI analysis m/z calculated $C_{15}H_{23}N_6O_5S_2$ $(M + H)^+$ 431.1; Gradient 1, 11 mL/min (Table 3-1); retention time 24.7 min, found 430.7; retention time 26.3 min, found 430.7.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-4-(isopropyldisulfanyl)butanoate (Hcp-A, 5b).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (29 mg, 0.036 mL, 0.18 mmol) were added to **4b** (38 mg, 0.0441 mmol) following the deprotection procedure given for **S4a**. HPLC/MALDI analysis m/z calculated $C_{17}H_{27}N_6O_5S_2$ $(M + H)^+$ 459.1; Gradient 1, 11 mL/min (Table 3-1); retention time 29.3 min, found 458.9; retention time 30.8 min, found 458.9.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-4-(*tert*-butyldisulfanyl)butanoate (Hcb-A, 5c).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (68 mg, 0.089

mL, 0.43 mmol) were added to **4c** (94.6 mg, 0.108 mmol) following the deprotection procedure given for **S4a**. HPLC/MALDI analysis m/z calculated $C_{18}H_{29}N_6O_5S_2$ $(M + H)^+$ 473.2; Gradient 1, 11 mL/min (Table 3-1); retention time 31.3 min, found 472.8; retention time 32.8 min, found 472.8.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)-tetrahydrofuran-3-yl 2-amino-4-(phenyldisulfanyl)butanoate (Hcf-A, 5d).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (65 mg, 0.084 mL, 0.409 mmol) were added to **4d** (91.5 mg, 0.1022 mmol) following the deprotection procedure given for **S4a**. HPLC/MALDI analysis m/z calculated $C_{20}H_{25}N_6O_5S_2$ $(M + H)^+$ 493.1; Gradient 3 (Table 3-1); retention time 32.1 min, found 493.1; retention time 33.2 min, found 493.1.

(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl) tetrahydrofuran-3-yl 2-amino-4-((2,2,2-trifluoroethyl)disulfanyl)butanoate (HcTf-A, 5e).

Trifluoroacetic acid (1 mL), tetrahydrofuran (1 mL) and TIPSH (67.1 mg, 0.087 mL, 0.421 mmol) were added to **4e** (94.9 mg, 0.105 mmol) following the deprotection procedure given for **S4a**. HPLC/MALDI analysis m/z calculated $C_{16}H_{22}F_3N_6O_5S_2$ $(M + H)^+$ 499.1; Gradient 1 at 11 mL/min (Table 3-1); retention time 29.1 min, found 499.1; retention time 30.6 min, found 499.1.

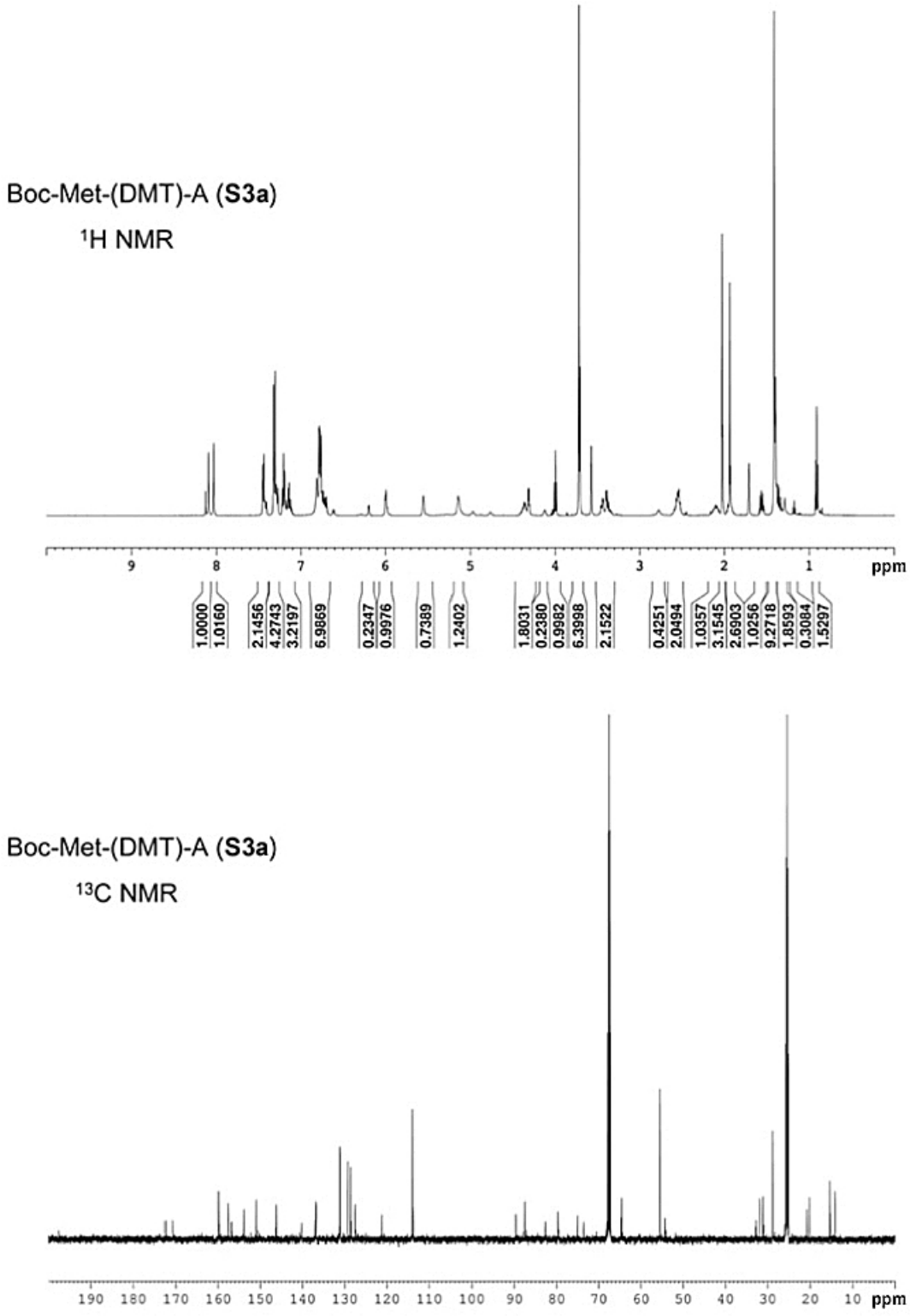


Figure 3-10. ^{1}H and ^{13}C NMR Characterization of Boc-Met-(DMT)-A (S3a)

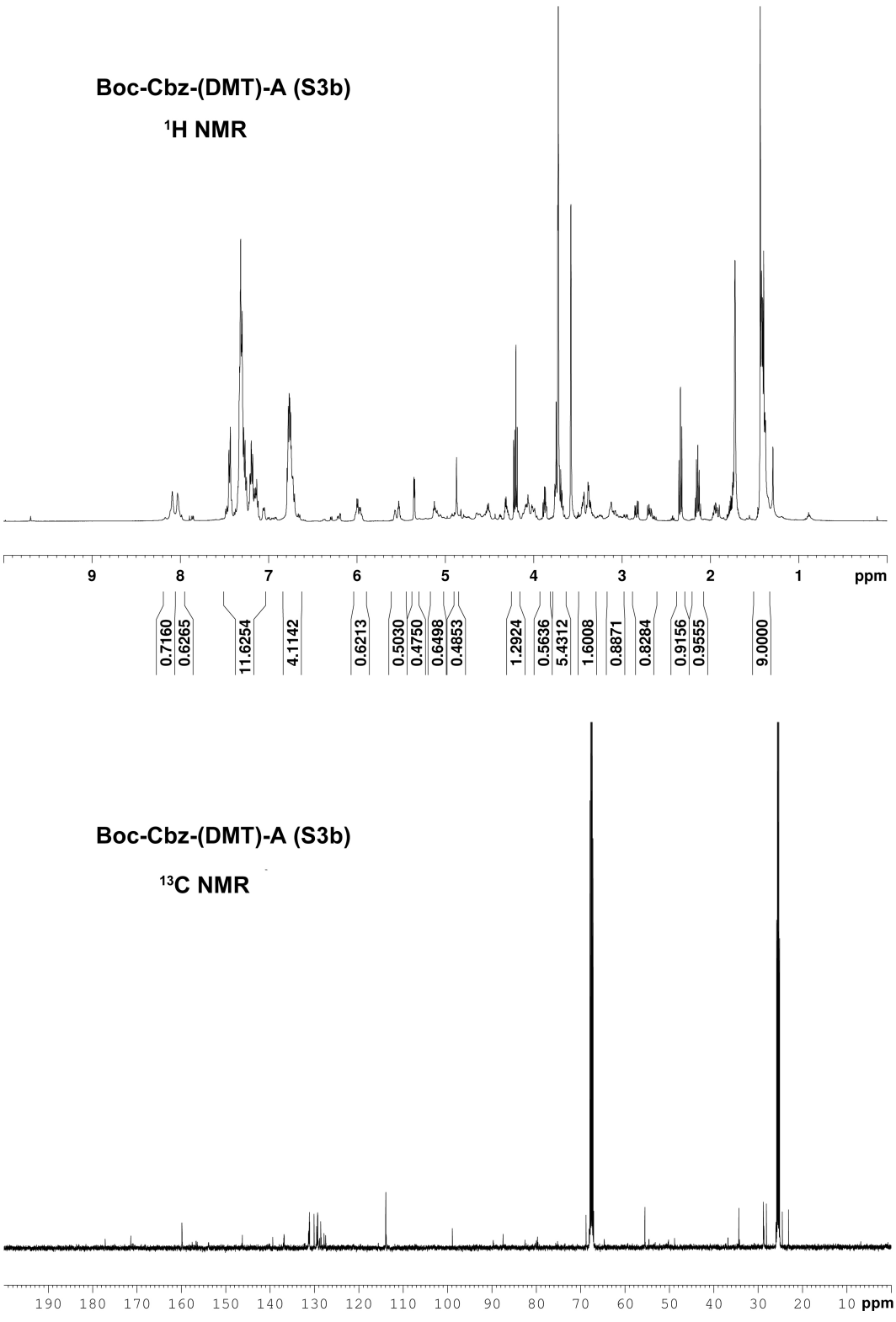


Figure 3-11. ^{1}H and ^{13}C NMR Characterization of Boc-Cys(Bz)-(DMT)-A (S3b)

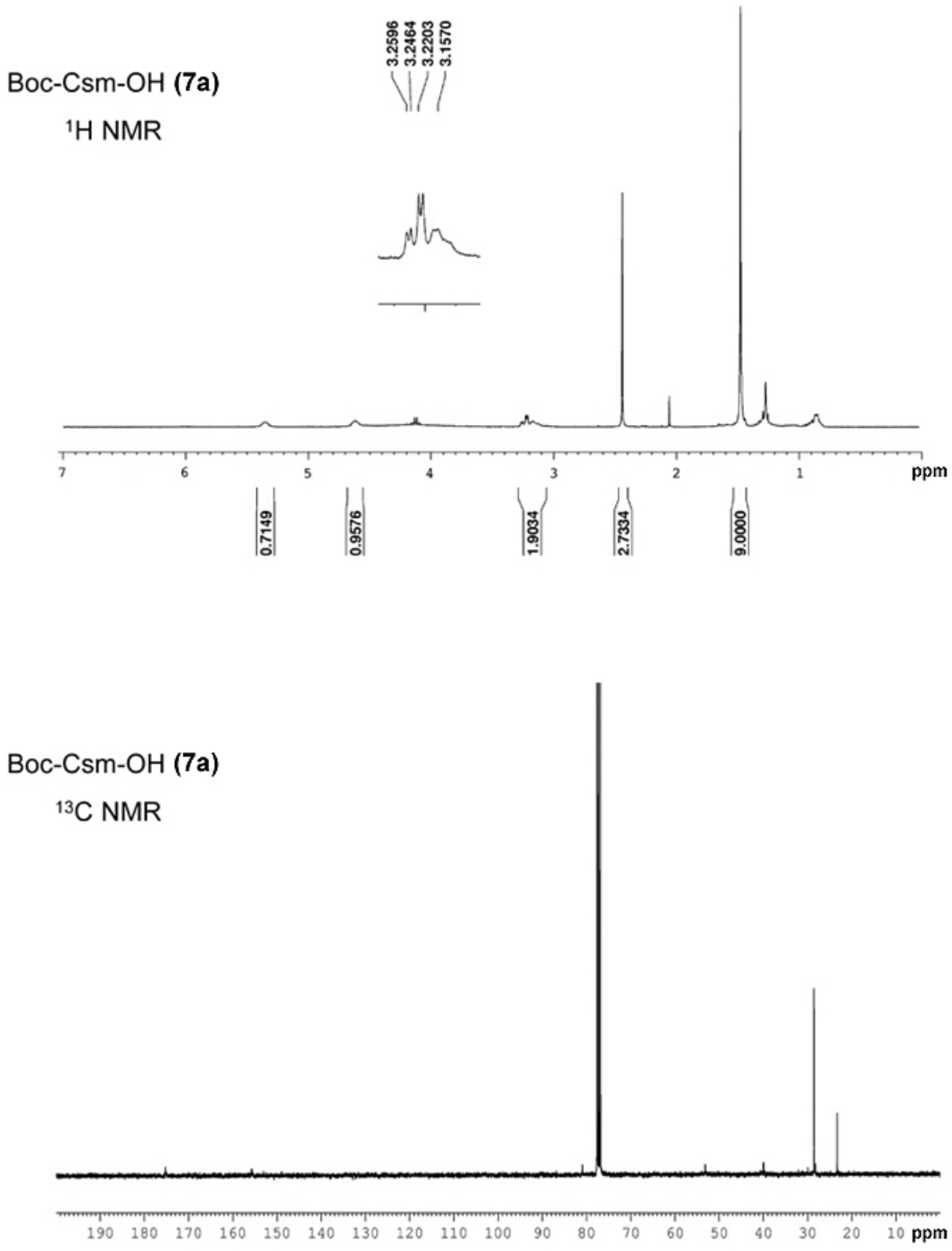


Figure 3-12. ^{1}H and ^{13}C NMR Characterization of Boc-Csm-OH (7a)

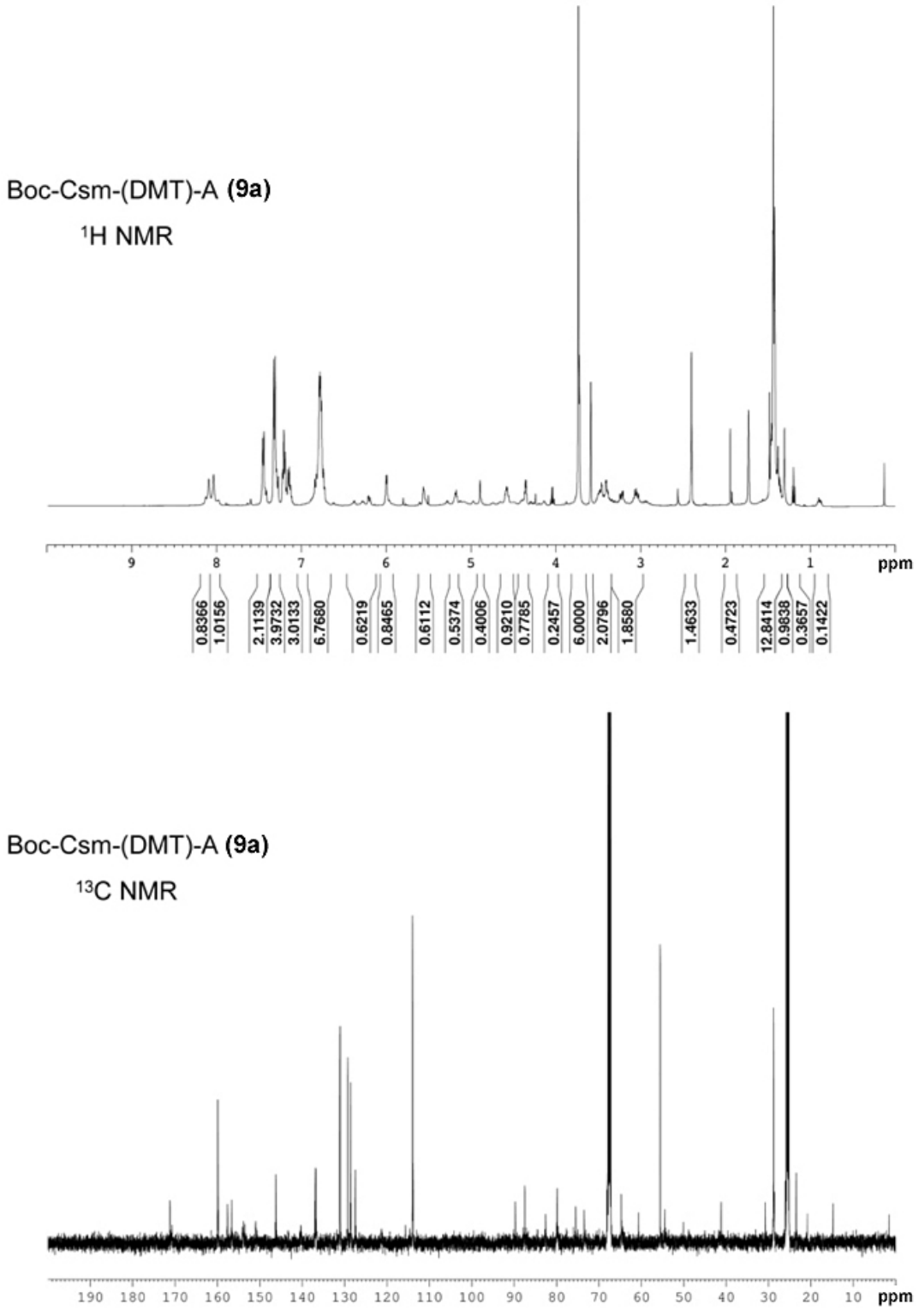


Figure 3-13. ^{1}H and ^{13}C NMR Characterization of Boc-Csm-(DMT)-A (9a)

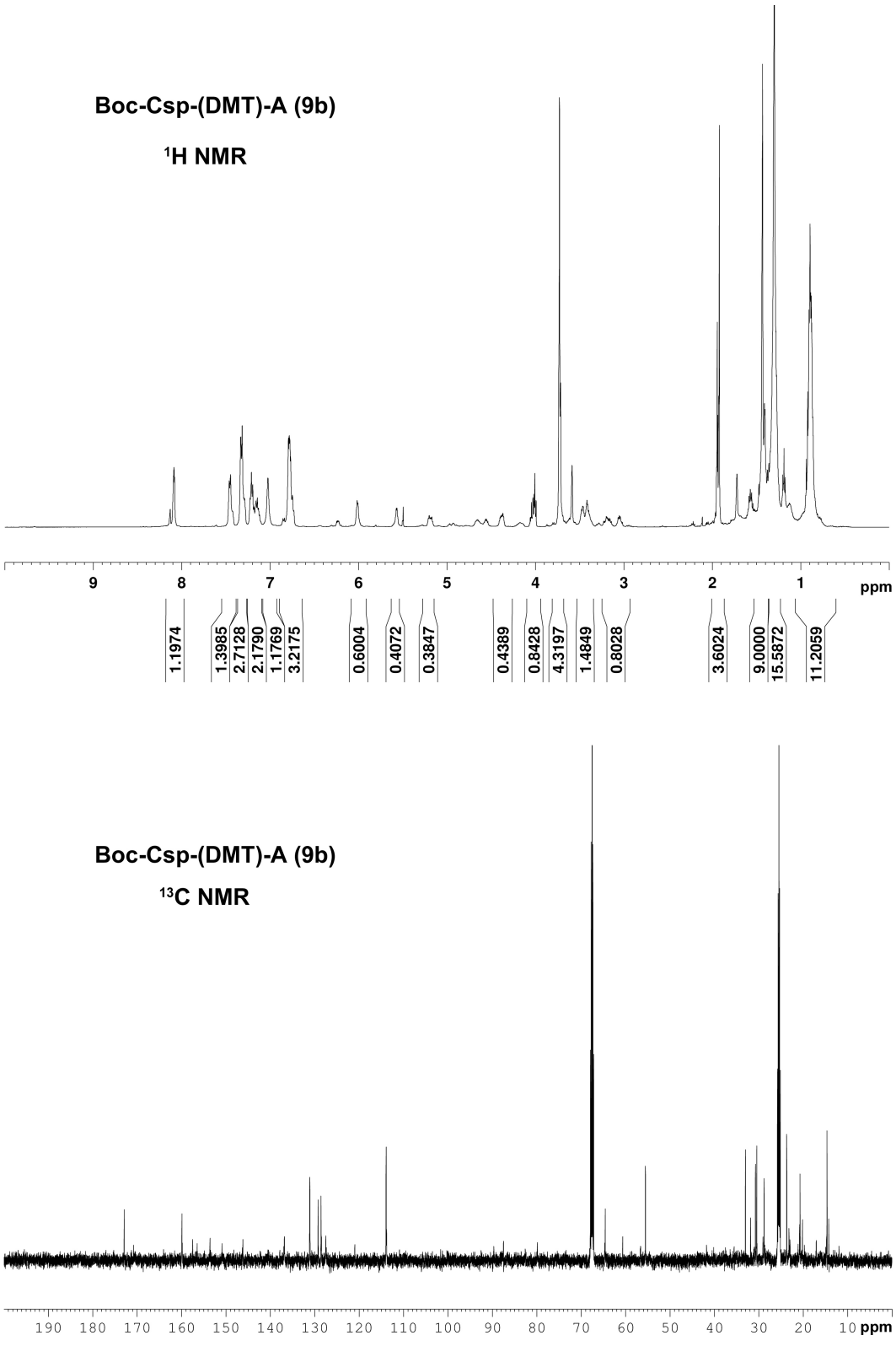


Figure 3-14. ^{1}H and ^{13}C NMR Characterization of Boc-Csp-(DMT)-A (9b)

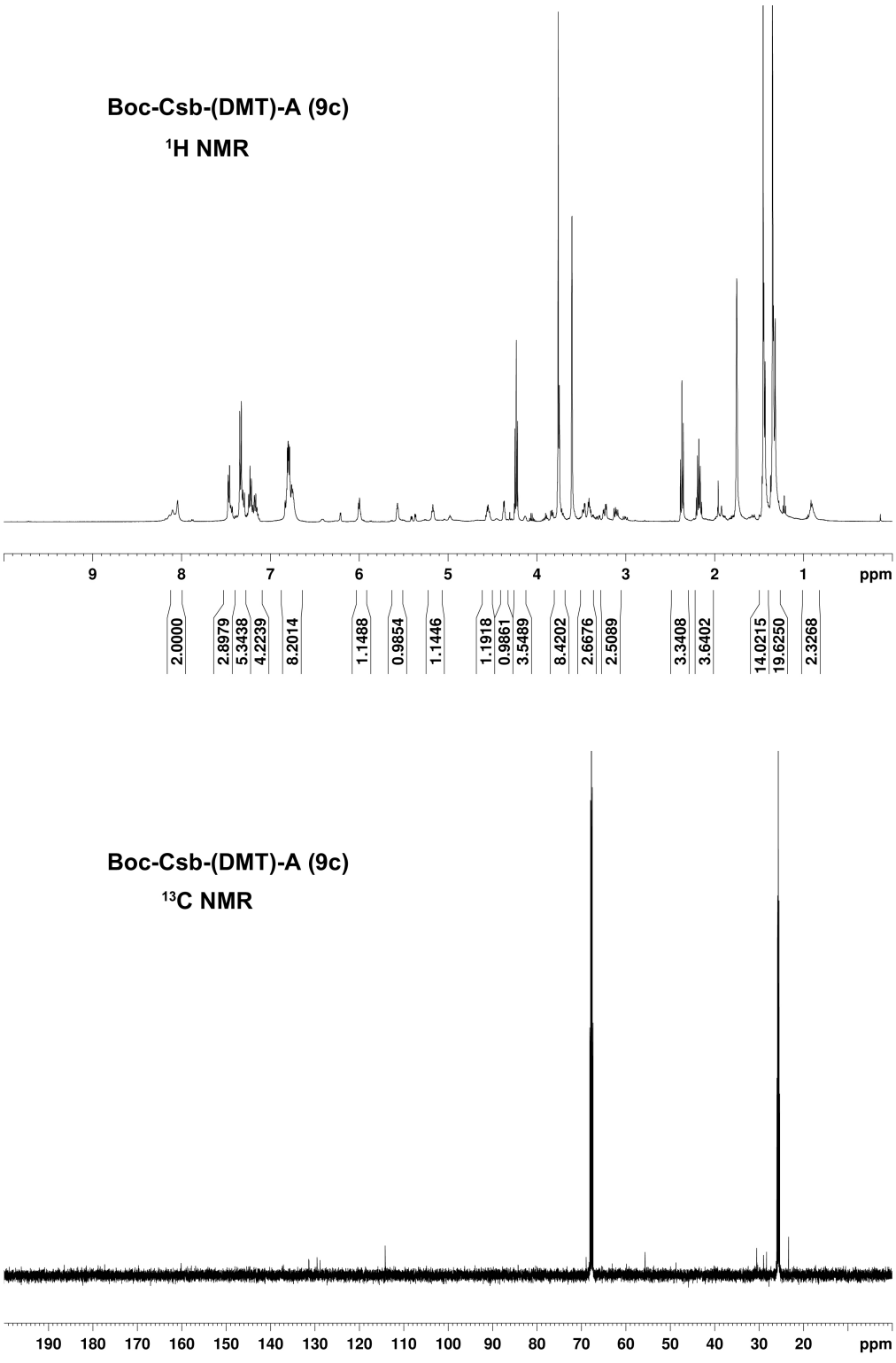


Figure 3-15. ^{1}H and ^{13}C NMR Characterization of Boc-Csb-(DMT)-A (9c)

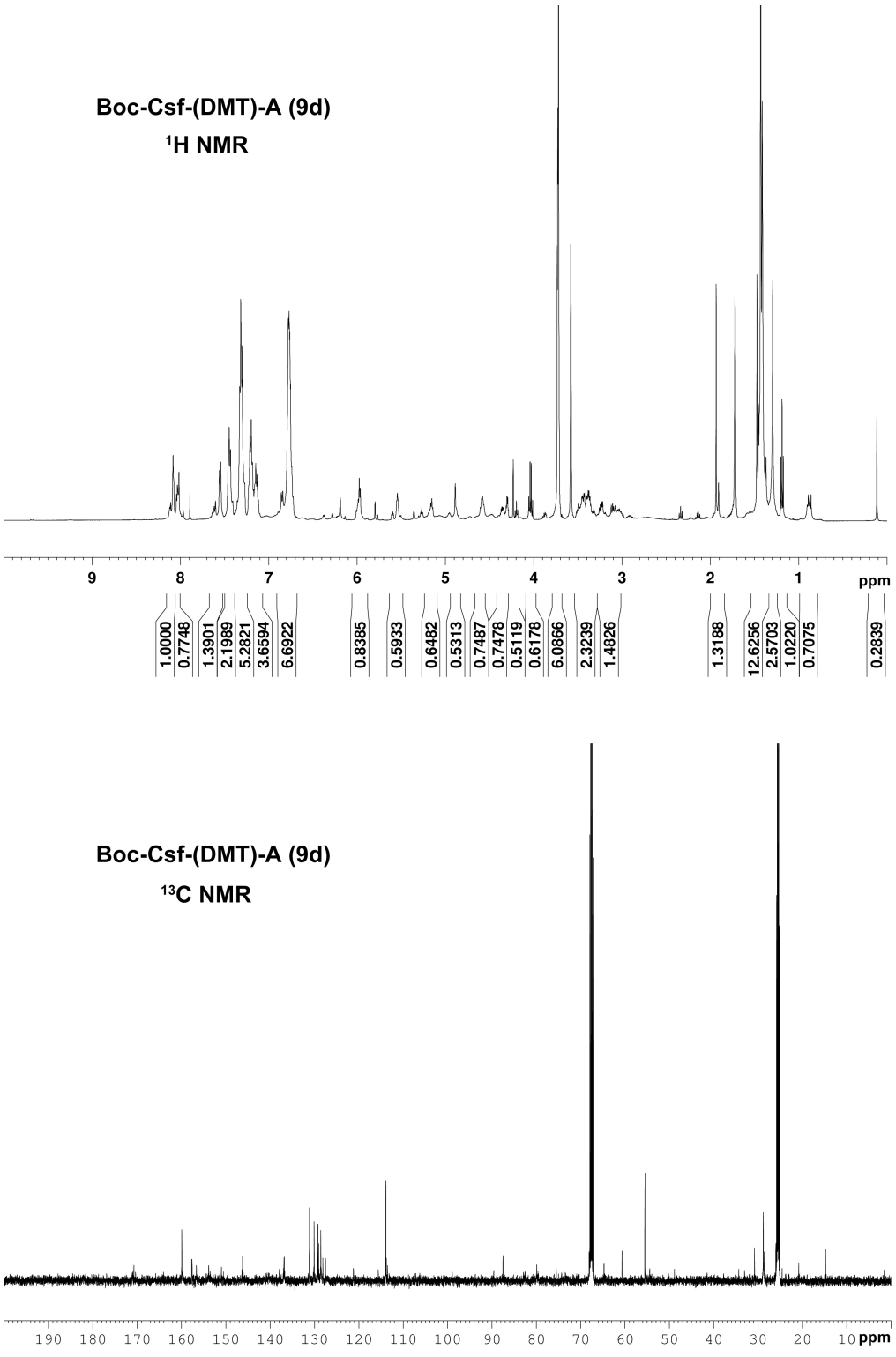


Figure 3-16. ^{1}H and ^{13}C NMR Characterization of Boc-Csf-(DMT)-A (9d)

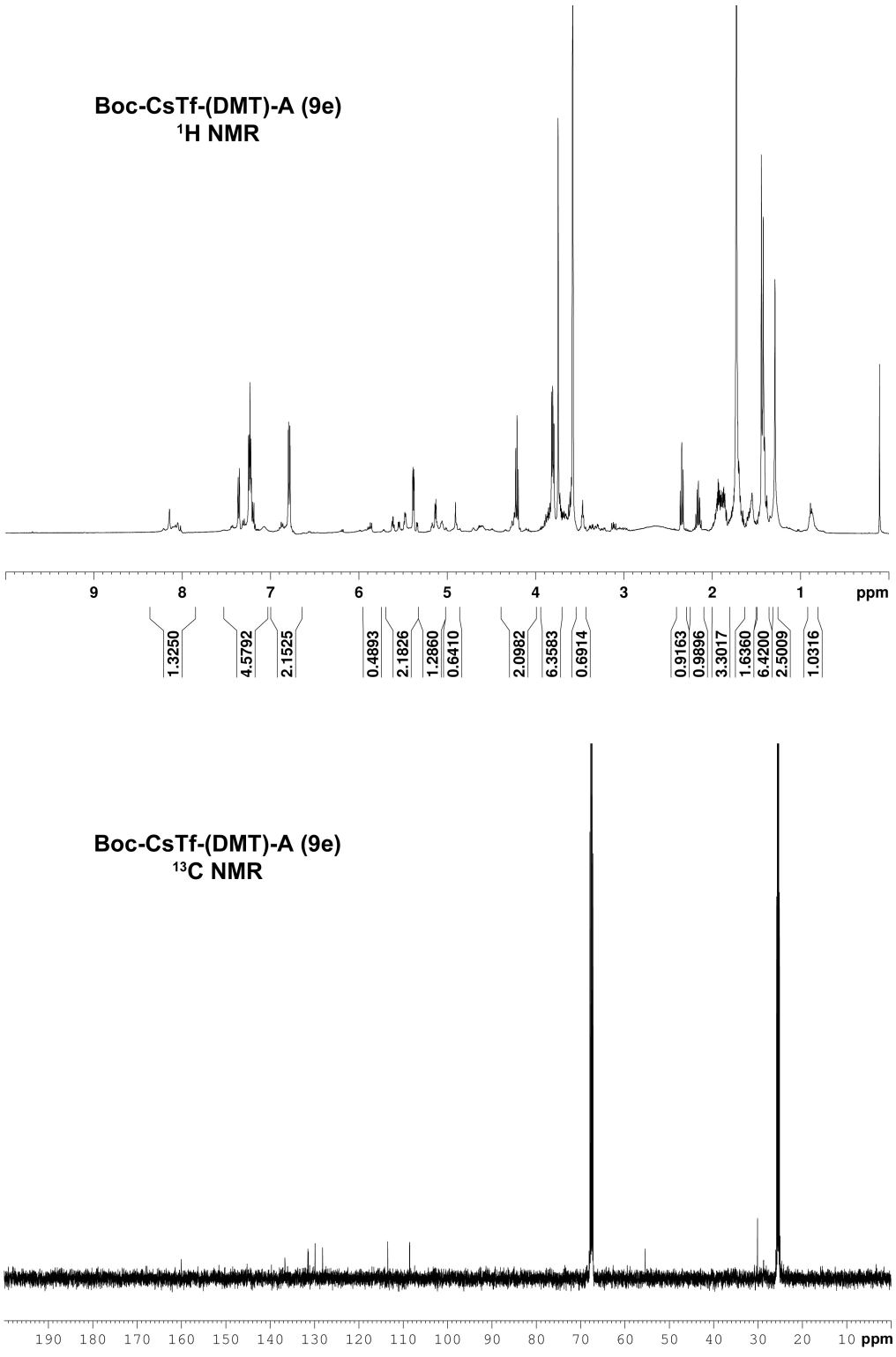


Figure 3-17. ^{1}H and ^{13}C NMR Characterization of Boc-CsTf-(DMT)-A (9e)

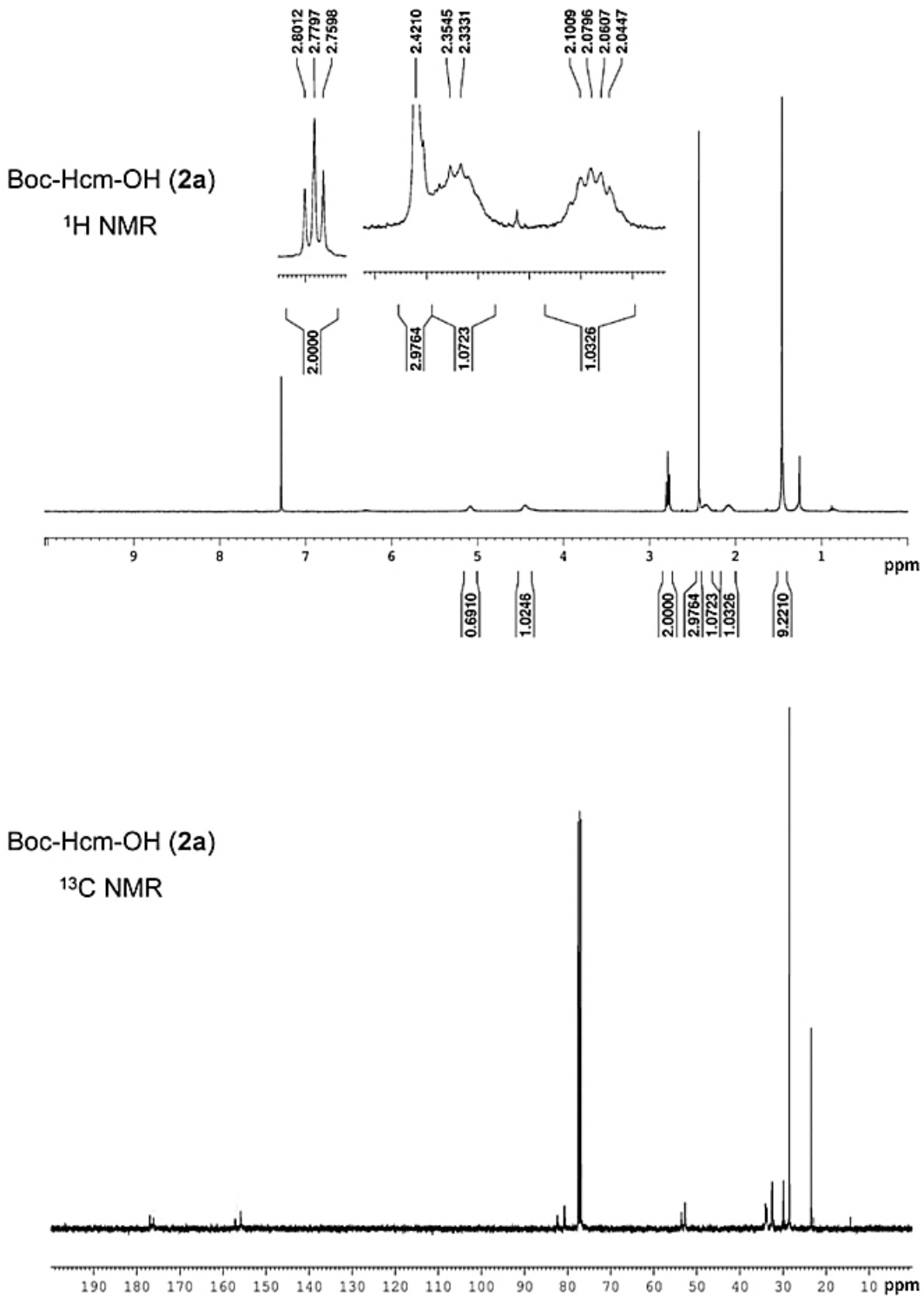
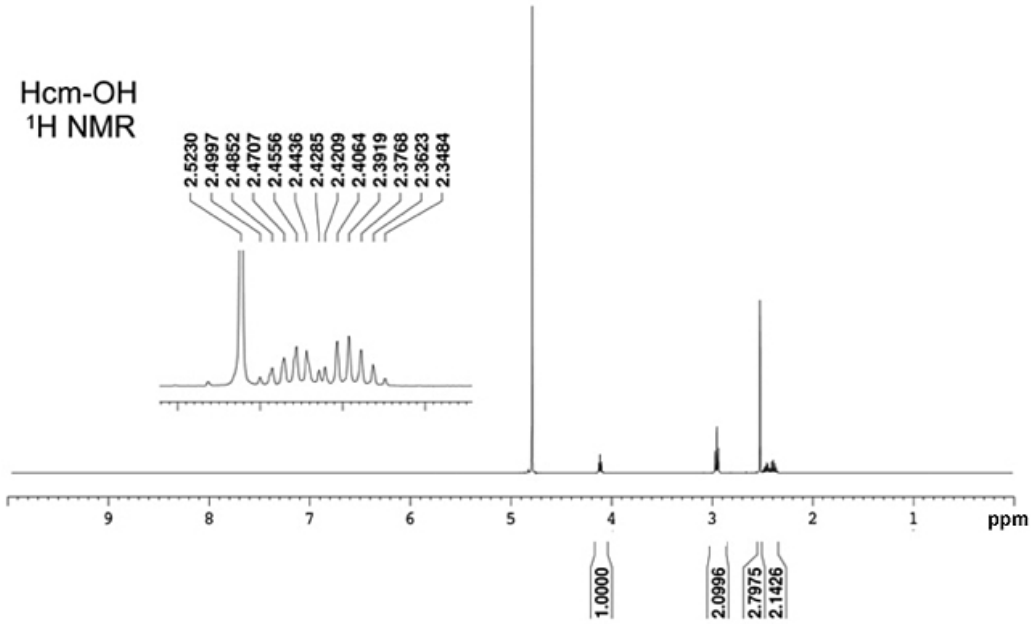


Figure 3-18. ^{1}H and ^{13}C NMR Characterization of Boc-Hcm-OH (2a)



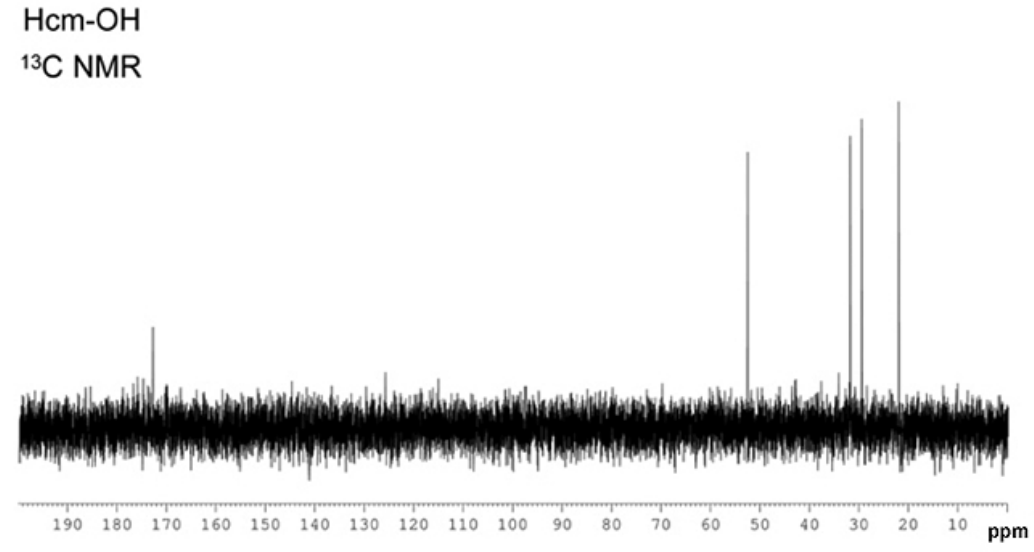


Figure 3-19. ^{1}H and ^{13}C NMR Characterization of Hcm-OH, Free Amino Acid

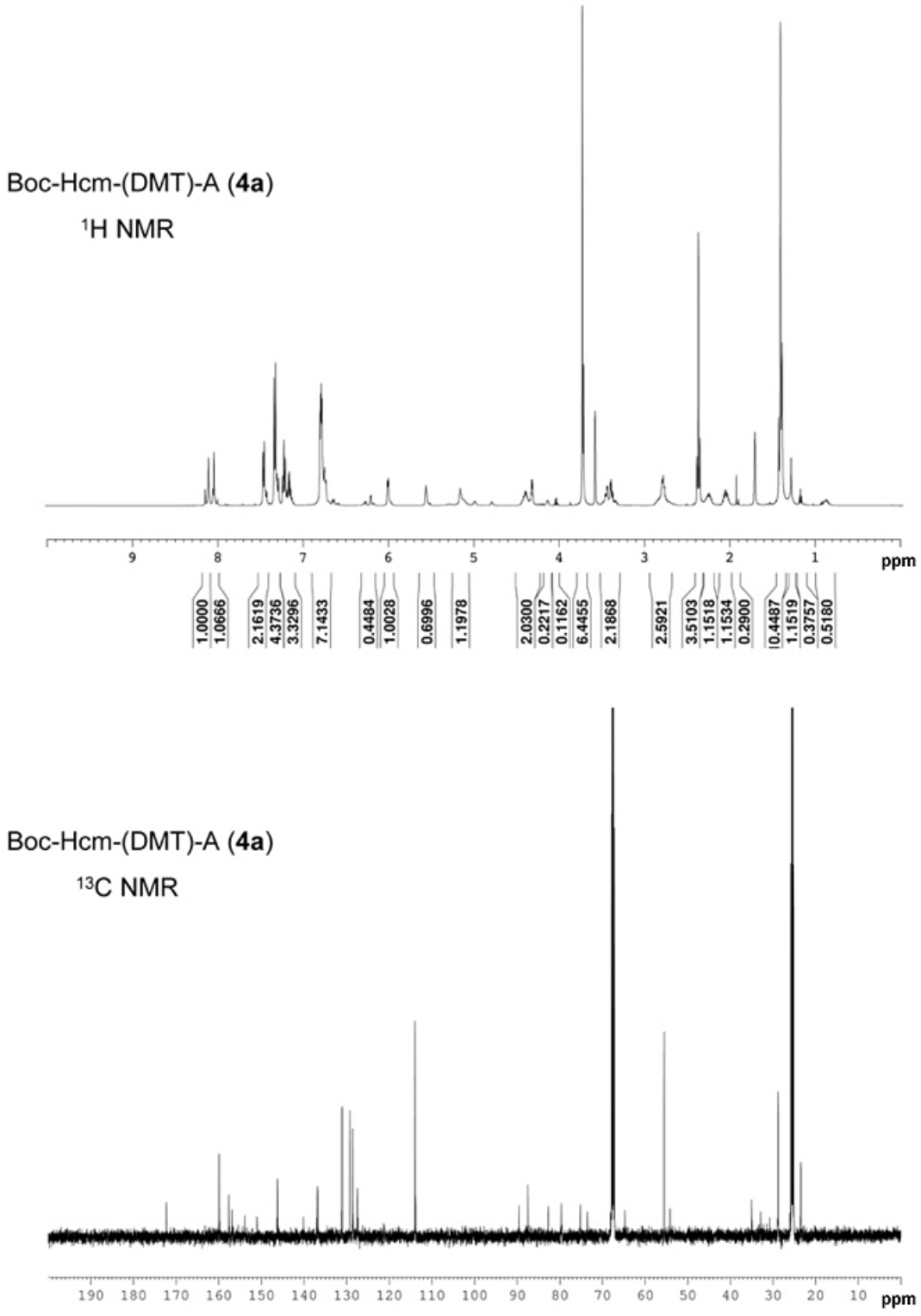


Figure 3-20. ^{1}H and ^{13}C NMR Characterization of Boc-Hcm-(DMT)-A (4a)

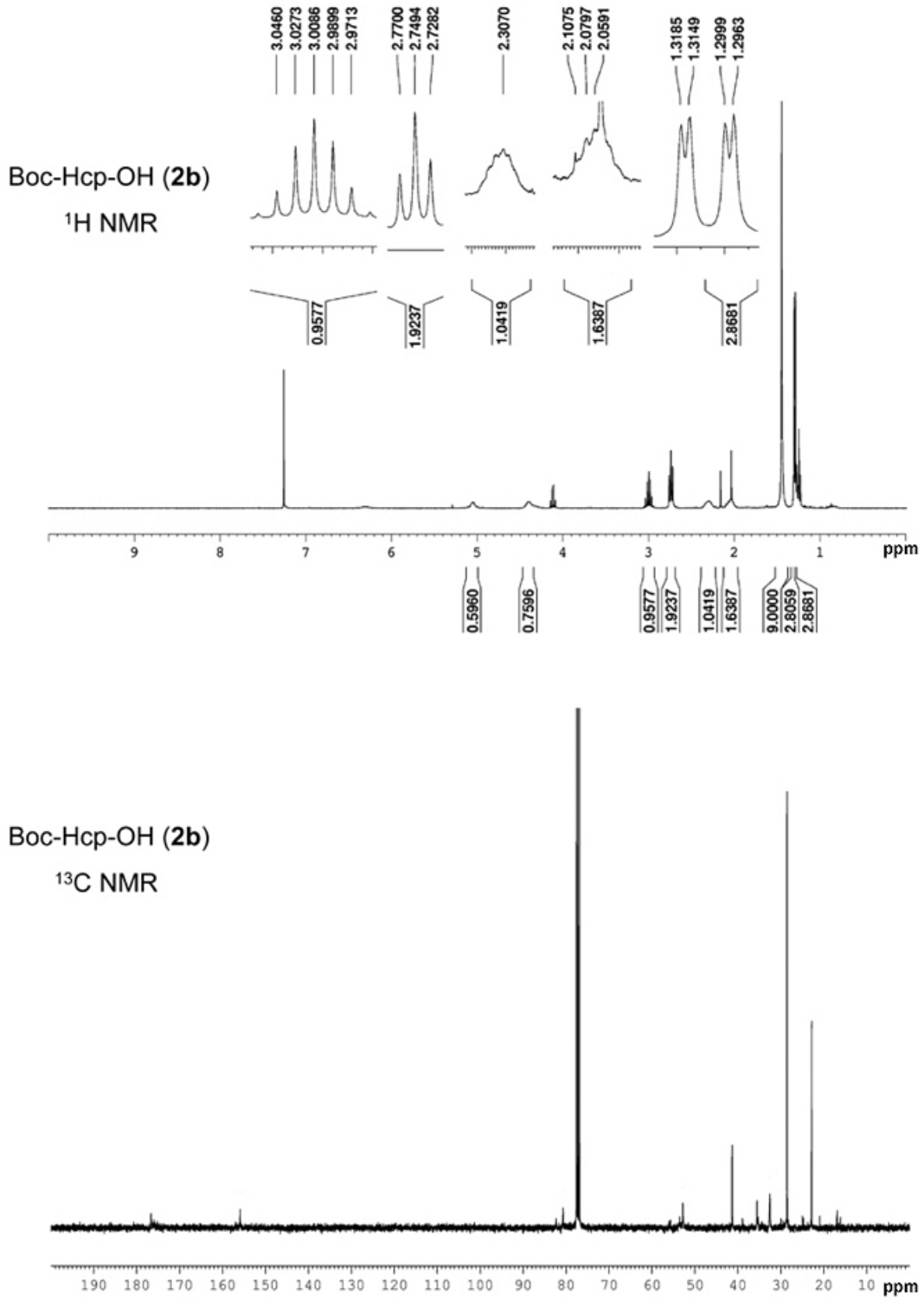


Figure 3-21. ^{1}H and ^{13}C NMR Characterization of Boc-Hcp-OH (2b)

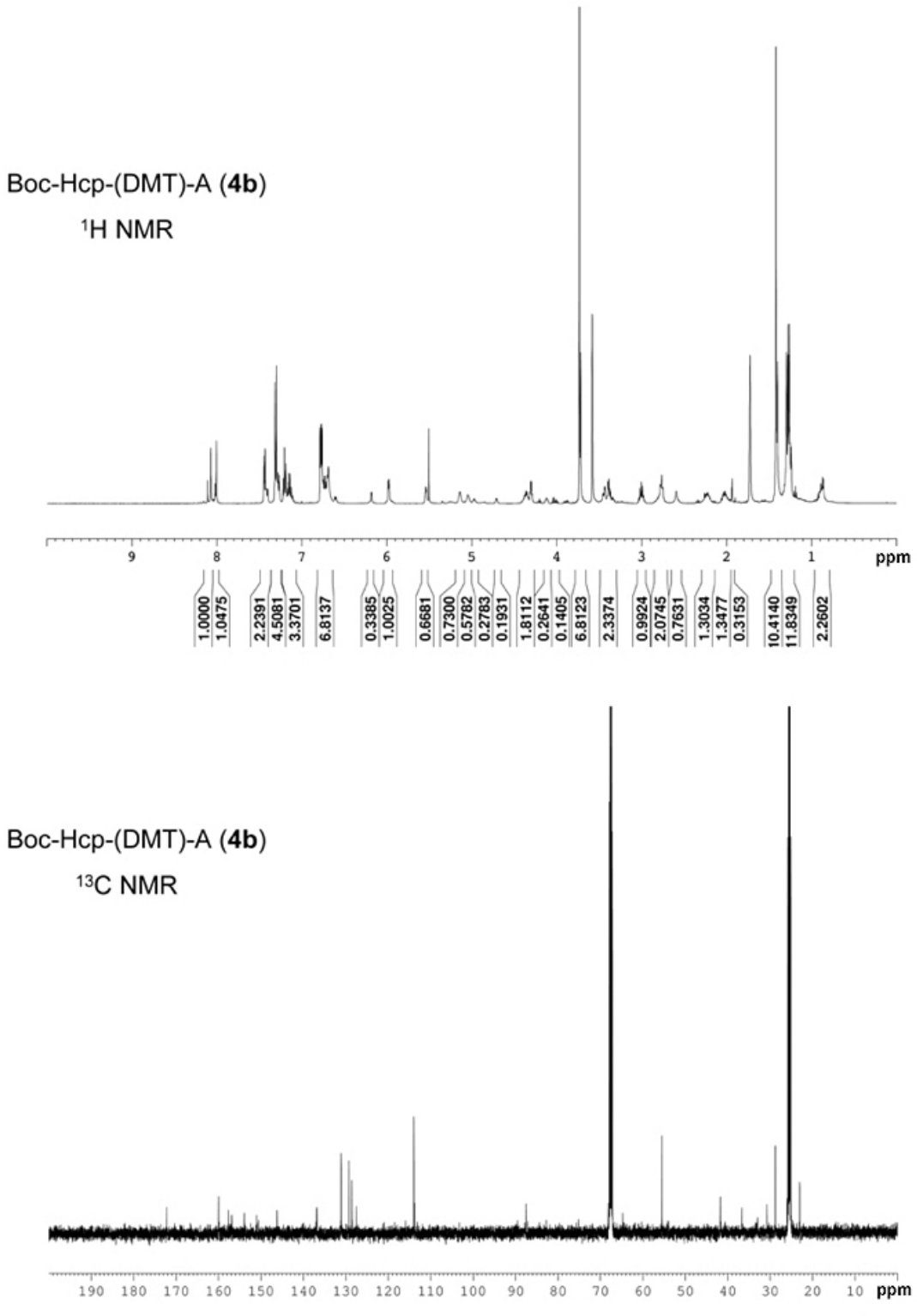


Figure 3-22. ^{1}H and ^{13}C NMR Characterization of Boc-Hcp-(DMT)-A (4b)

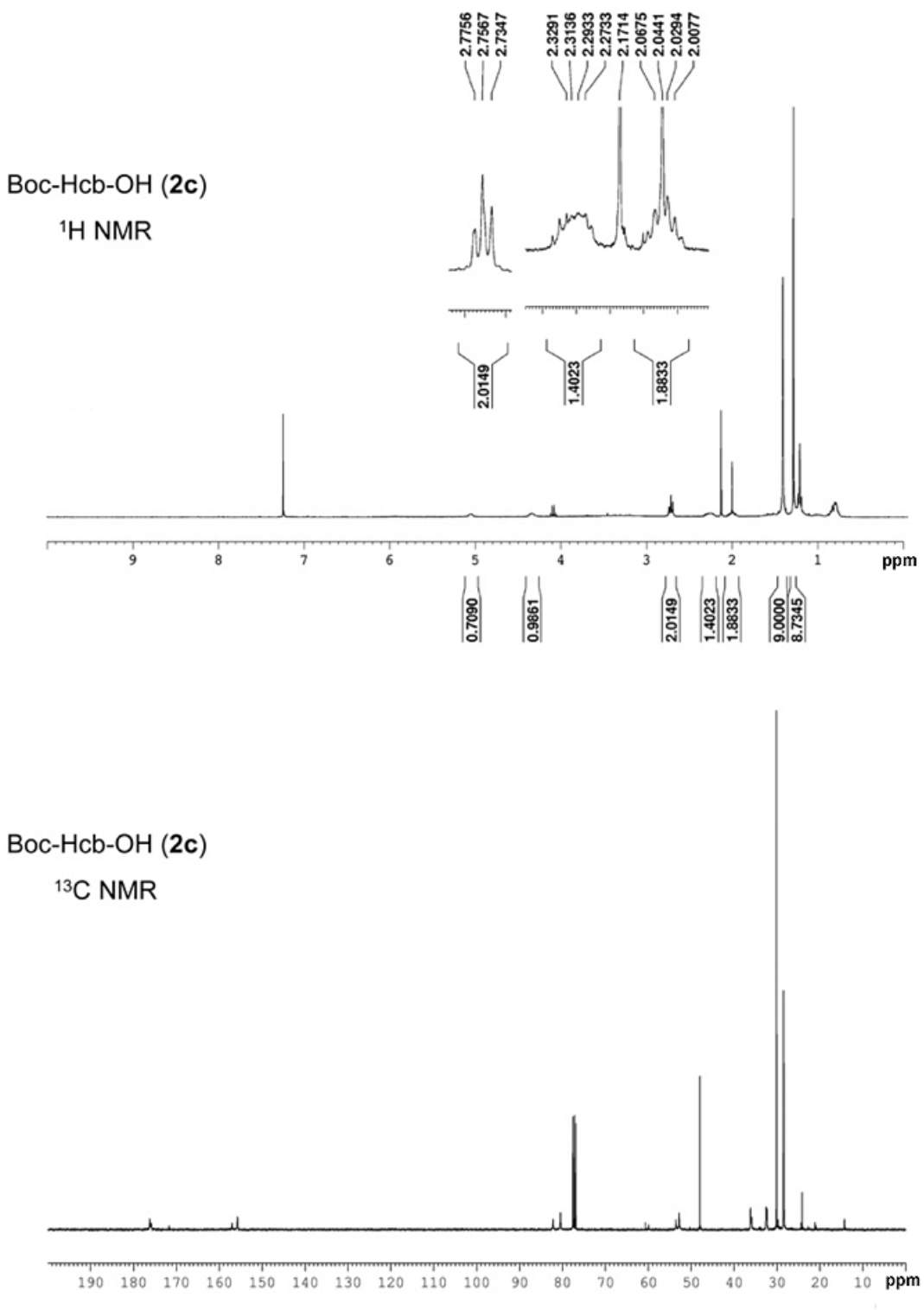


Figure 3-23. ^{1}H and ^{13}C NMR Characterization of Boc-Hcb-OH (2c)

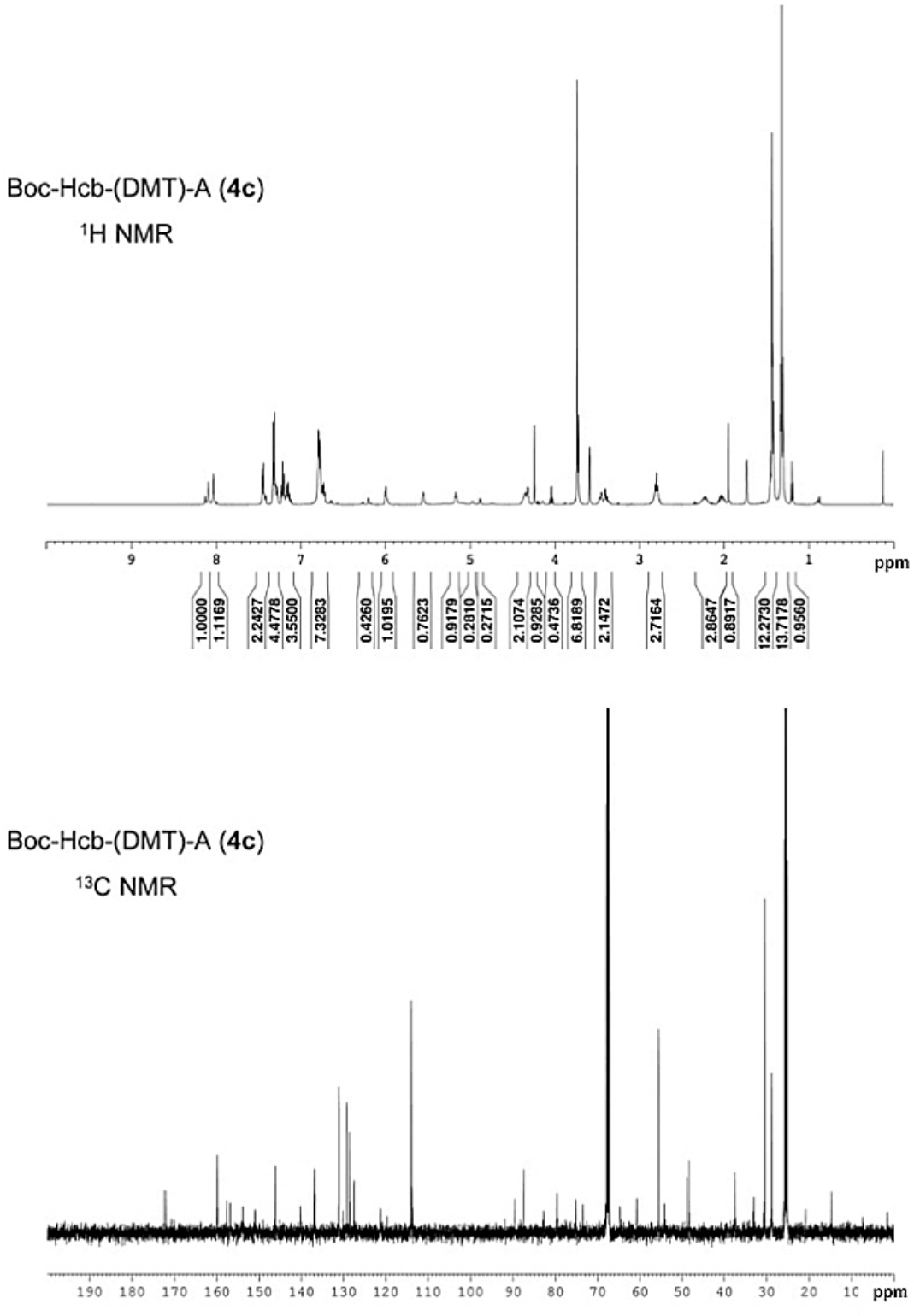


Figure 3-24. ^{1}H and ^{13}C NMR Characterization of Boc-Hcb-(DMT)-A (4c)

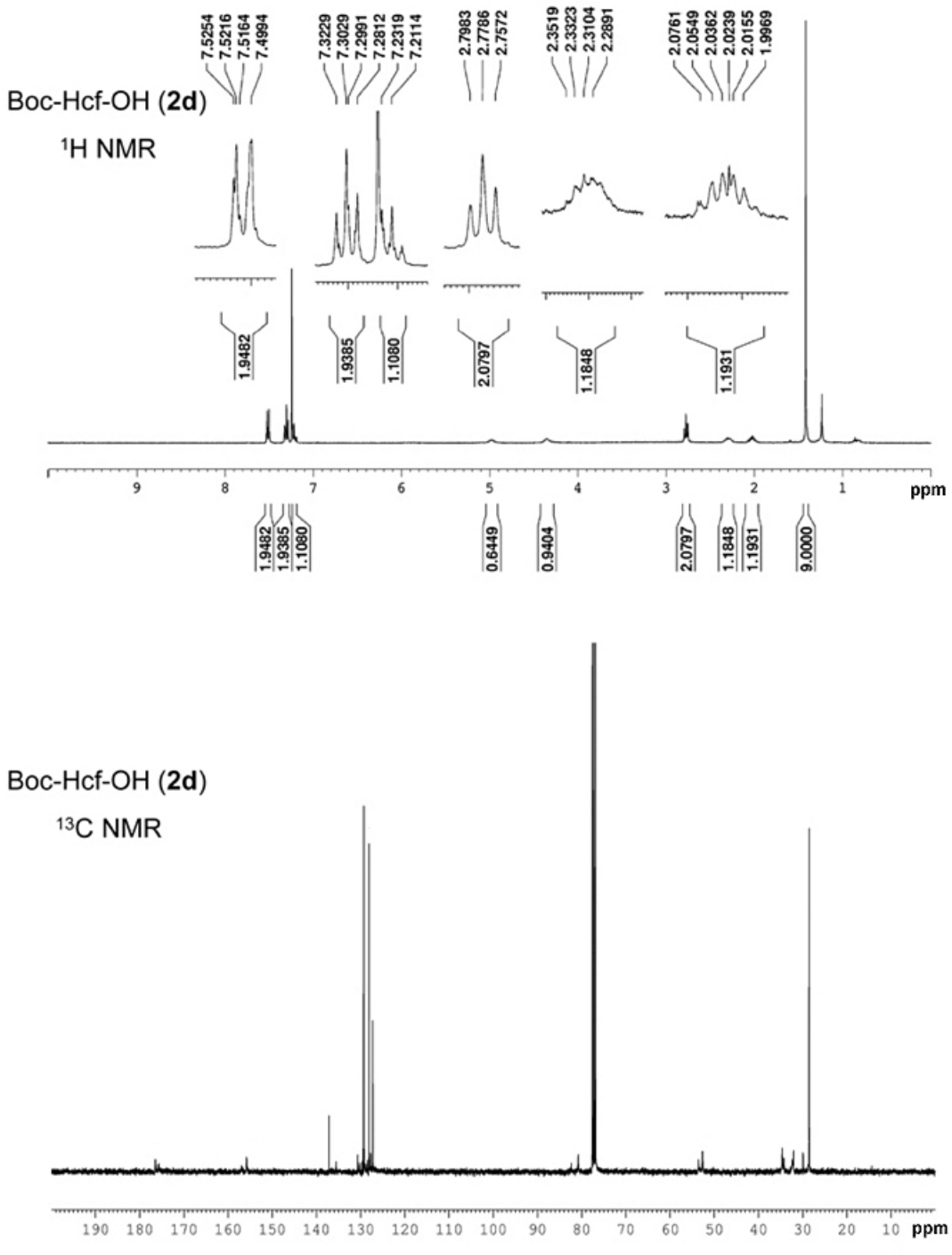


Figure 3-25. ^{1}H and ^{13}C NMR Characterization of Boc-Hcf-OH (2d)

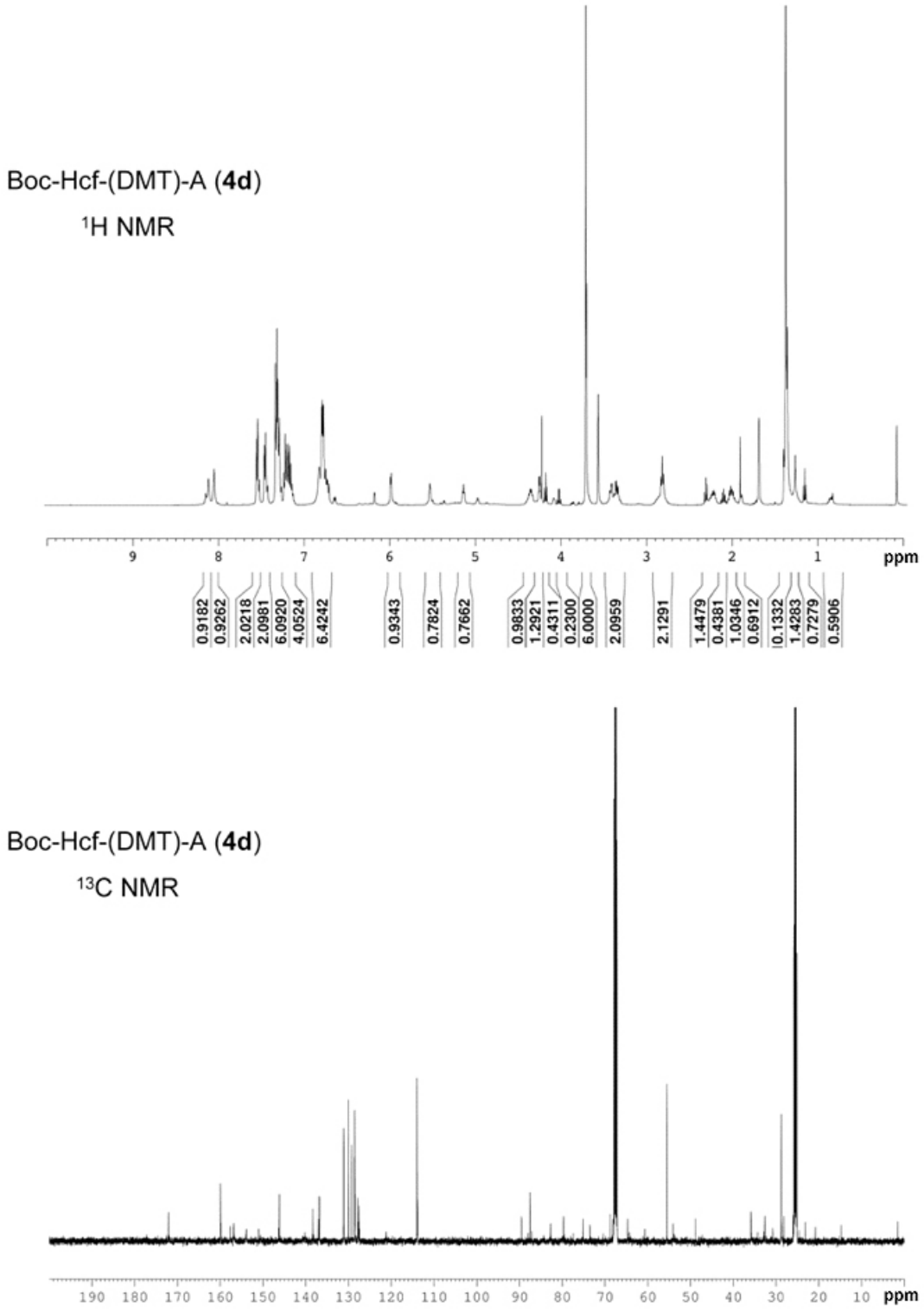


Figure 3-26. ^{1}H and ^{13}C NMR Characterization of Boc-Hcf-(DMT)-A (4d)

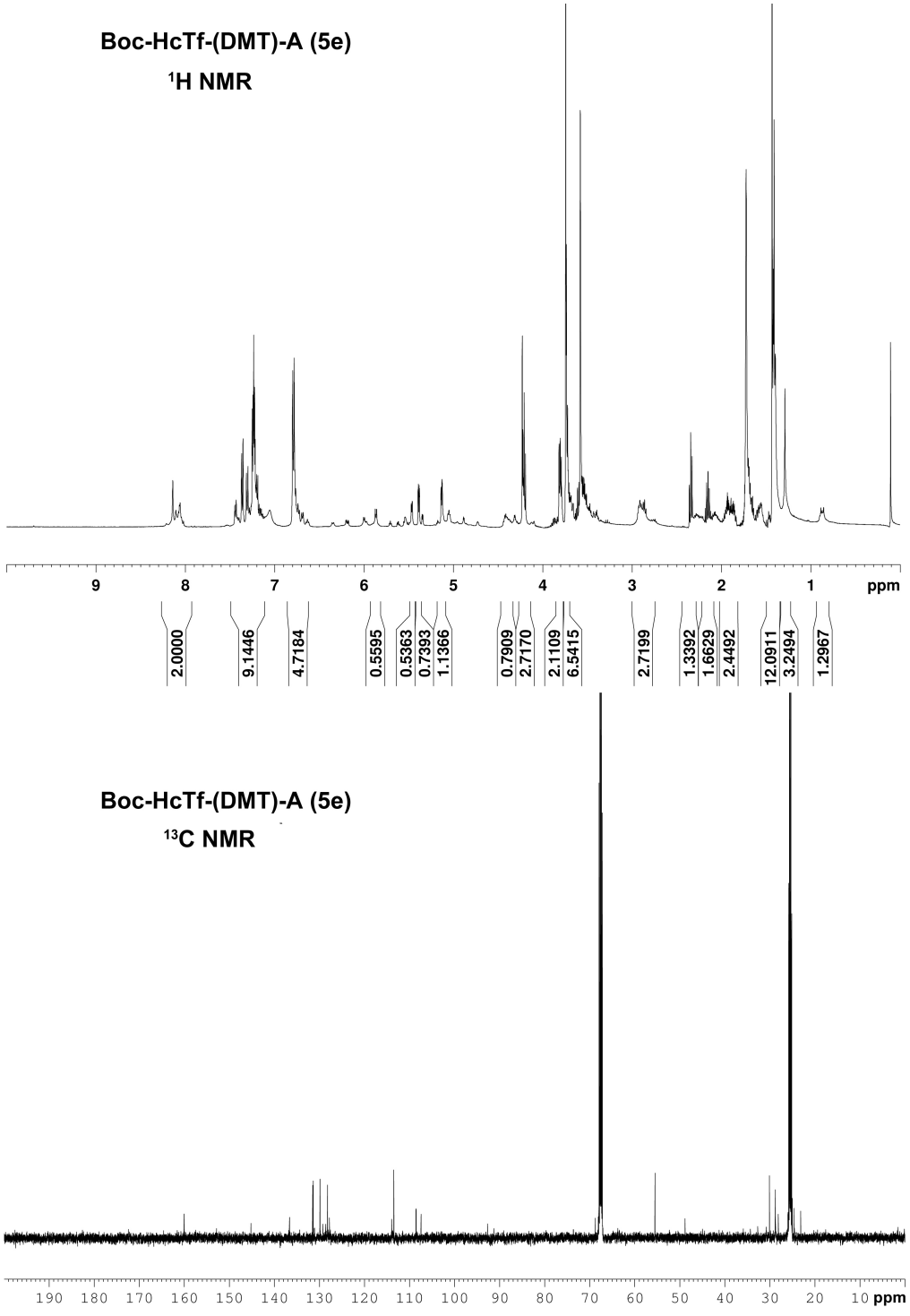


Figure 3-27. ^{1}H and ^{13}C NMR Characterization of Boc-HcTf-(DMT)-A (4e)

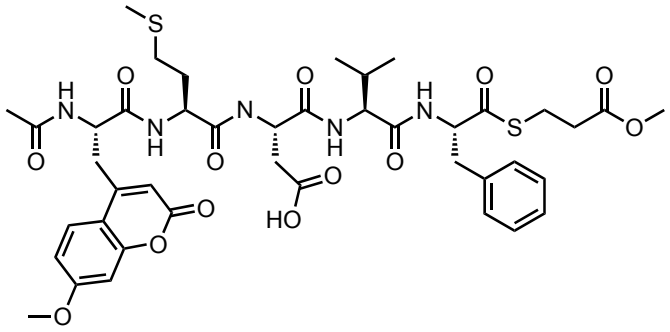
3.3.3 – Peptide synthesis

All peptides were synthesized using a manual, Fmoc-solid phase procedure on 2-chlorotrityl (ClTrt) resin as described previously.[173] Protected $\alpha S_{1\text{-}4}$ peptides Ac-MetAsp(*t*-Bu)ValPhe-OH (**S6**) and Ac-McmMetAsp(*t*-Bu)ValPhe-OH (**S7**), where Mcm is β-(7-methoxy-coumarin-4-yl)-alanine, were cleaved by treatment with acetic acid /trifluoroethanol/dichloromethane (1/1/3, v/v) for 1 h. After the resin was filtered, solvent was removed *in vacuo*. Ac-MetAspValPheHcsLys (**S8**) and Ac-MetAspValPheMetLys (**S9**) were cleaved with TFA/triisopropylsilane/water (95/2.5/2.5, v/v) for 1 h, and purified by C18 reverse phase HPLC using Gradient 4a (**S8**) and Gradient 4b (**S9**).

3.3.4 – Thioesterification of Ac-MetAspValPhe-SR (11) and Ac-McmMetAspValPhe-SR (S10)

100 μmol of protected **S6** or **S7** was dissolved in 10 mL of DMF in an ice bath. 10 equiv. each of PyBOP and triethylamine were added to the peptide solution. Then, 20 equiv. of methyl mercaptopropionate were added and the solution was allowed to stir overnight. Solvent was removed *in vacuo*. Pouring water onto the residue caused immediate precipitation of a white solid. The precipitate was collected by centrifugation, and washed again with water. The crude protected peptide thioester was dissolved in TFA/triisopropylsilane/water (95/2.5/2.5, v/v) and allowed to stir for 1 h. After removal of the solvent *in vacuo*, pouring $E_{t2}O$ onto the residue gave a precipitate that was collected by centrifuge, and washed with Et_2O (3 times). Deprotected thioesters **11** or **S10** (see structure below) were then purified by C18 reverse phase semi-prep HPLC

using Gradient 5.



S10

Table 3-1. Solvent Gradients Used for Peptide Purification and Analysis

Gradient #	Time (min)	Buffer A (%)	Gradient #	Time (min)	Buffer A (%)
1	0:00	99	**2**	0:00	0
(rate in text)	10:00	99	15 mL/min	5:00	0
	40:00	65		35:00	15
	45:00	0		40:00	100
	50:00	0		45:00	100
	55:00	99		50:00	0
	60:00	99			
3	0:00	99	**4a**	0:00	71
12 mL/min	10:00	99	14 mL/min	5:00	71
	40:00	65		35:00	65
	50:00	0		40:00	0
	55:00	0		45:00	0
	65:00	99		50:00	71
	70:00	99			
4b	0:00	71	**5**	0:00	71
11 mL/min	5:00	71	4 mL/min	5:00	71
	35:00	65		35:00	65
	40:00	0		40:00	0
	45:00	0		45:00	0
	50:00	71		50:00	71
6	0:00	99	**7**	0:00	99
1 mL/min	5:00	99	4 mL/min	5:00	99
	10:00	70		10:00	85
	15:00	60		42:00	55
	20:00	0		46:00	0
	25:00	0		49:00	0
	27:00	99		52:00	99
	30:00	99		55:00	99

*3.3.5 – MetRS and Met*RS Expression and Purification*

E. coli MetRS and Met*RS were obtained from David Tirell. Plasmids were transformed into *E. coli* BL21 (DE3) and grown to OD_{600} 0.6 in LB media. Protein expression was induced with 1.0 mM isopropyl β–D–thiogalactoside and grown for 18 h at 25 °C. The cells were harvested at 5000 x g for 15 min and the resulting pellet was resuspended in a lysis buffer (300 mM KCl, 50 mM Tris pH 8.0) and included protease inhibitor cocktail, 1 mM PMSF, and 10 units/mL DNAse1–Grade II) and sonicated. Soluble protein was obtained by centrifugation for 20 minutes at 30,000 x g, 4 °C. Collected soluble protein was gently shaken for 1 h on ice with Ni-NTA resin. Protein was purified by rinsing with lysis buffer followed by washing with wash buffer (300 mM KCl, 50 mM Tris pH 8.0, and 15 mM imidazole). Protein was eluted with 4 x 1 mL Elution Buffer A (300 mM KCl, 50 mM Tris pH 8.0, and 50 mM imidazole) followed by 4 x 1 mL Elution Buffer B (300 mM KCl, 50 mM Tris pH 8.0, and 250 mM imidazole). SDS-PAGE analysis was performed to analyze the purity of elution fractions. Collected protein fractions were dialyzed against storage buffer (50 mM Tris pH 7.5 in 50 % glycerol) at 4 °C. Protein was stored in 0.5 mL aliquots at -80 °C.

3.3.6 – Chemoenzymatic LysAlaAcm Ligation Assay

Each ligation reaction, 125 μL total volume, contained the following reagents: aminoacyl adenosine donor (1 mM), recombinant His_{10}-tagged *E. coli* aminoacyl transferase (2.26 μM), and LysAlaAcm (100 μM) in the AaT Ligation Buffer (50 mM HEPES pH 8.0, 150 mM KCl, 10 mM $MgCl_2$). The reaction mixtures were incubated at 37 °C for four hours. At the second, third, and fourth hour of the reaction, an additional 1

mM dose of aminoacyl adenosine donor was added to the reaction mixture. The reaction was quenched with 1 % acetic acid at 4 h. One reaction material modification was made for Hcm analog only, in which we prepared AaT in transferase buffer without the presence of β-mercaptoethanol (50 mM Tris, 30 % glycerol, 120 mM $(NH_4)_2SO_4$). The proteins were extracted from the reactions *via* acetone precipitation. The reactions were precipitated using 4X reaction volume of acetone and cooled at - 20 °C for 1 hour. Next, the reactions were centrifuged at 13,200 rpm at 4 °C for 20 min to separate the reaction from precipitated protein. The supernatant was transferred to fresh 1.5 mL centrifuge tubes and allowed for acetone evaporation overnight at room temperature for all non-disulfide analogs. For all disulfide analogs, the reactions were rotary evaporated immediately to remove acetone in order to minimize disulfide reduction due to air exposure. After acetone evaporation, the supernatant was dried in a Speedvac (Savant, Thermo Scientific, Fisher Inc.) for 30 min to remove residual acetone. The resulting reaction volume was dissolved up to 1.2 mL using MilliQ water and analyzed by HPLC (Gradient 6) to determine ligation yield by integration of separated peak intensities monitored at 325 nm. Collected HPLC fractions were characterized through MALDI MS analysis. Reactions were performed with 3 trials on at least two protein preparations for all successful reactions.

3.3.7 – Fully Enzymatic LysAlaAcm Ligation Assay

Each ligation reaction, 125 µL total volume, contained the following reagents: amino acid (1 mM), ATP (2.5 mM), total *E. coli* tRNA (2 µg/µL), recombinant His-tagged *E. coli* MetRS or Met*RS, recombinant His_{10}-tagged *E. coli* aminoacyl transferase

(2.26 μM), and LysAlaAcm (100 μM) in the AaT Ligation Buffer (50 mM HEPES pH 8.0, 150 mM KCl, 10 mM $MgCl_2$). The reaction mixtures were incubated at 37 °C for 30 minutes and quenched with 1 % acetic acid. Acetone precipitation and subsequent analysis were performed as described for the Chemoenzymatic LysAlaAcm Ligation Assay.

3.3.8 – AaT Ligation Reaction Analysis by HPLC

All HPLC analyses of LysAlaAcm ligations were monitored using a Waters Symmetry Shield C18 reversed phase column (Gradient 6). Peptide retention times monitored at 325 nm (Acm absorption) are in Table 3-3 and MALDI MS analyses, confirming peptide identity, are in Table 3-2.

Table 3-2. Calculated and Observed Peptide and Protein Masses

Peptide	Calculated $(M+H)^+$	Observed $(M+H)^+$	Calculated $(M+Na)^+$	Observed $(M+Na)^+$
MetLysAlaAcm[a]	506.243	506.114	528.225	528.1
CsmLysAlaAcm[a]	524.200	523.892	546.182	545.893
CbzLysAlaAcm[a]	568.259	---	590.241	---
CspLysAlaAcm[a]	552.231	552.436	574.213	---
CsbLysAlaAcm[a]	566.247	566.400	588.229	---
CsfLysAlaAcm[a]	586.215	---	608.198	---
CsTfLysAlaAcm[a]	592.187	---	614.169	---
HcsLysAlaAcm[a*]	492.611	---	514.209	---
HcmLysAlaAcm[a] **(14)**	538.215	538.162	560.197	560.167
HcpLysAlaAcm[a]	566.247	566.393	588.228	588.357
HcbLysAlaAcm[a]	580.262	579.923	602.224	601.924
HcfLysAlaAcm[a]	600.231	---	622.213	---
HcTfLysAlaAcm[a]	606.203	---	628.185	---
Ac-MetAspValPhe-SR[a] **(11)**	655.247	---	677.229	677.302
Ac-MetAspValPheHcsLys[a] **(S8)**	798.352	798.600	820.334	820.591
Ac-MetAspValPheMetLys[a] **(S9)**	812.368	812.449	834.350	834.437
Ac-McmMetAspValPhe-SR[a] **(S10)**	900.31	---	922.31	922.25
Ac-MetAspValPheHcsLysAlaAcm[a] (**15**)	1026.44	1026.59	1048.42	1048.56
Ac-MetAspValPheMetLysAlaAcm[a] (**16**)	1040.46	1040.24	1062.44	---
$\alpha S_{6\text{-}140}$[b] **(18)**	13, 837.3	13, 837.7	13, 859.3	---
Hcm-$\alpha S_{6\text{-}140}$[b] **(19)**	14, 000.6	14, 001.0	14, 022.6	---
AcMcm-αS-Hcs_5[b] **(20a)**	14, 733.4	14, 735.1	14, 755.4	---
AcMcm-αS[b] **(21a)**	14, 747.4	14, 748.4	14, 769.4	---
Ac-αS-Hcs_5[b] **(20b)**	14, 489.2	14, 490.0	14, 511.2	---
Ac-αS[b] **(21b)**	14, 503.2	14, 502.7	14, 525.2	---

- [a]Exact mass, [b]Molecular weight
- HcsLysAlaAcm observed as a byproduct of transfer of **5a**, **5b**, or **5c** in the presence of BME.

Table 3-3. HPLC Analysis of LysAlaAcm Transfer Reactions

Peptide	Time (min)	Conditions	% Yield
LysAlaAcm	11.3		NA
PheLysAlaAcm	12.5	S1	99.5 ± 0.1
PheLysAlaAcm	12.5	S1, 0.05 M Gdn•HCl	96.0 ± 0.1
PheLysAlaAcm	12.5	S1, 0.10 M Gdn•HCl	93.4 ± 0.1
PheLysAlaAcm	12.5	S1, 0.20 M Gdn•HCl	79.2 ± 5.0
PheLysAlaAcm	12.5	S1, 0.30 M Gdn•HCl	71.2 ± 4.0
PheLysAlaAcm	12.5	S1, 0.40 M Gdn•HCl	62.0 ± 3.8
PheLysAlaAcm	12.5	S1, 0.50 M Gdn•HCl	54.5 ± 0.4
PheLysAlaAcm	12.5	S1, 1.00 M Gdn•HCl	25.4 ± 2.4
PheLysAlaAcm	12.5	S1, 1.50 M Gdn•HCl	10.6 ± 0.4
PheLysAlaAcm	12.5	S1, 2.00 M Gdn•HCl	3.0 ± 0.2
PheLysAlaAcm	12.5	S1, 0.1 % Triton	95.0 ± 3.8
PheLysAlaAcm	12.5	S1, 0.5 % Triton	97.8 ± 0.1
PheLysAlaAcm	12.5	S1, 1.0 % Triton	96.5 ± 2.4
PheLysAlaAcm	12.5	S1, 2.0 % Triton	97.4 ± 1.5
PheLysAlaAcm	12.5	S1, 5.0 % Triton	98.4 ± 0.5
MetLysAlaAcm	12.1	S4a	74.6 ± 1.3
MetLysAlaAcm	12.1	Met, MetRS	99.9 ± 0.1
MetLysAlaAcm	12.1	Met, 1 mg/mL Met*RS, 0.20 M Gdn•HCl	74.2 ± 8.7
MetLysAlaAcm	12.1	Met, 1 mg/mL Met*RS, 0.50 M Gdn•HCl	30.1 ± 1.3
MetLysAlaAcm	12.1	Met, 1mg/mL Met*RS, 5.0 % Triton	92.0 ± 9.6
CbzLysAlaAcm	NA	S4b	0
CsmLysAlaAcm	12.3	5*a	20.9 ± 0.5
CspLysAlaAcm	13.0	5*b	68.4± 0.2
CsbLysAlaAcm	13.3	5*c	41.3 ± 0.7
CsfLysAlaAcm	NA	5*d	0
CsTfLysAlaAcm	NA	5*e	0
HcmLysAlaAcm	12.7	5a	21.2 ± 0.7
HcmLysAlaAcm	12.7	5a, 0 mM BME	36.5 ± 0.1
HcmLysAlaAcm	12.7	5a, 1 mg/mL MetRS	99.9 ± 0.1
HcmLysAlaAcm	12.7	5a, 0.1mg/mL MetRS, 30 min	48.8 ± 4.8
HcmLysAlaAcm	12.7	5a, 1 mg/mL Met*RS	99.9 ± 0.1
HcmLysAlaAcm	12.7	5a, 0.1mg/mL Met*RS, 30 min	63.8 ± 3.1
HcmLysAlaAcm	12.7	5a, 0.1mg/mL Met*RS	99.9 ± 0.1
HcmLysAlaAcm	12.7	5a, 1 mg/mL Met*RS, 5.0 % Triton	99.4 ± 0.1
HcmLysAlaAcm	12.7	5a, 1 mg/mL Met*RS, 0.20 M Gdn•HCl	76.7 ± 8.1
HcmLysAlaAcm	12.7	5a, 1 mg/mL Met*RS, 0.50 M Gdn•HCl	15.3 ± 8.3
HcpLysAlaAcm	13.5	5b	14.6 ± 0.3
HcbLysAlaAcm	13.7	5c	2.7 ± 0.2
HcfLysAlaAcm	NA	5d	0
HcTfLysAlaAcm	NA	5e	0

3.3.9 – Ligation of Ac-MetAspValPhe-SR (11) and HcmLysAlaAcm (14)

Ligation buffer was prepared as follows: 200 mM sodium phosphate dibasic, 40 mM TCEP, 1 % v/v thiophenol and 1 % v/v benzylmercaptan, pH 7.2. Ligation buffer was prepared fresh and sparged with argon for 10 – 20 minutes. 50 uL of the ligation buffer was added to 50 uL of a 2 mM stock of HcmLysAlaAcm. The solution was then added to 0.33 μmol of Ac-MetAspValPhe-SR and the reaction mixture was sparged for an additional 15 minutes. The reaction was mixed overnight using a micro-stir bar. The reaction was quenched the following day with a 10-fold amount of 0.1 % trifluoroacetic acid and analyzed by HPLC (Gradient 7). Peptide retention times were monitored at 215 nm (peptide bond absorption) and MALDI MS analyses, confirming peptide identity, are in Table 3-2.

3.3.10 – Construction of Recombinant $\alpha S_{6\text{-}140}$ Expression Plasmid

We have previously described the cloning of the αS gene from a pT7-7 vector into the pET-16b prokaryotic expression vector, which contains an N-terminal His tag followed by a Factor Xa recognition site (IEGR) (Figure 3-28 Ia).[2] QuikChange® mutagenesis was used to delete residues H_{-1} and M_1 (Figure 3-28 Ib) followed by a D2K mutation (Figure 3-28 Ic) to yield the $\alpha S_{2\text{-}140}K_2$ plasmid. Residues K_2-M_5 were deleted (Figure 3-28 Id) from the $\alpha S_{2\text{-}140}K_2$ plasmid using QuickChange® mutagenesis to give the $\alpha S_{6\text{-}140}$ plasmid. All mutagenesis steps were confirmed by DNA sequencing using the T7 promoter primer. Plasmid constructed by John Warner.

I. Construction of $\alpha S_{6\text{-}140}$ Plasmid DNA

(A)	5' - ATC GAA GGT CGT **CAT ATG** GAT GTA TTC ATG AAA GGA CTT TCA AAG GCC -3' I E G R H_{-1} M_1 D_2 V_3 F_4 M_5 K_6 G_7 L_8 S_9 K_{10} A_{11}
(B)	5' - ATC GAA GGT CGT GAT GTA TTC ATG AAA GGA CTT TCA AAG GCC -3' I E G R D_2 V_3 F_4 M_5 K_6 G_7 L_8 S_9 K_{10} A_{11}
(C)	5' - ATC GAA GGT CGT **AAA** GTA TTC ATG AAA GGA CTT TCA AAG GCC -3' I E G R D_2 V_3 F_4 M_5 K_6 G_7 L_8 S_9 K_{10} A_{11}
(D)	5' - ATC GAA GGT CGT AAA GGA CTT TCA AAG GCC -3' I E G R K_6 G_7 L_8 S_9 K_{10} A_{11}

II. DNA Oligomers Used for QuickChange® Mutagenesis

(A) Deletion of H_{-1} and M_1	Forward 5' – GGCCATATCGAAGGTCGTGATGTATTCATGAAAGGA – 3'
	Reverse 5' – TCCTTTCATGAATACATCACGACCTTCGATATGGCC – 3'
(B) Mutation of D_2K	Forward 5' – GCGGCCATATCGAAGGTCGTAAAGTATTCATGAAAGGACTTTC – 3'
	Reverse 5' – GAAAGTCCTTTCATGAATACTTTACGACCTTCGATATGGCCGC – 3'
(C) Deletion of K_2 – M_5	Forward 5' – GGCCATATCGAAGGTCGTAAAGGACTTTCAAAGGCCAAGG – 3'
	Reverse 5' – CCTTGGCCTTTGAAAGTCCTTTACGACCTTCGATATGGCC – 3'

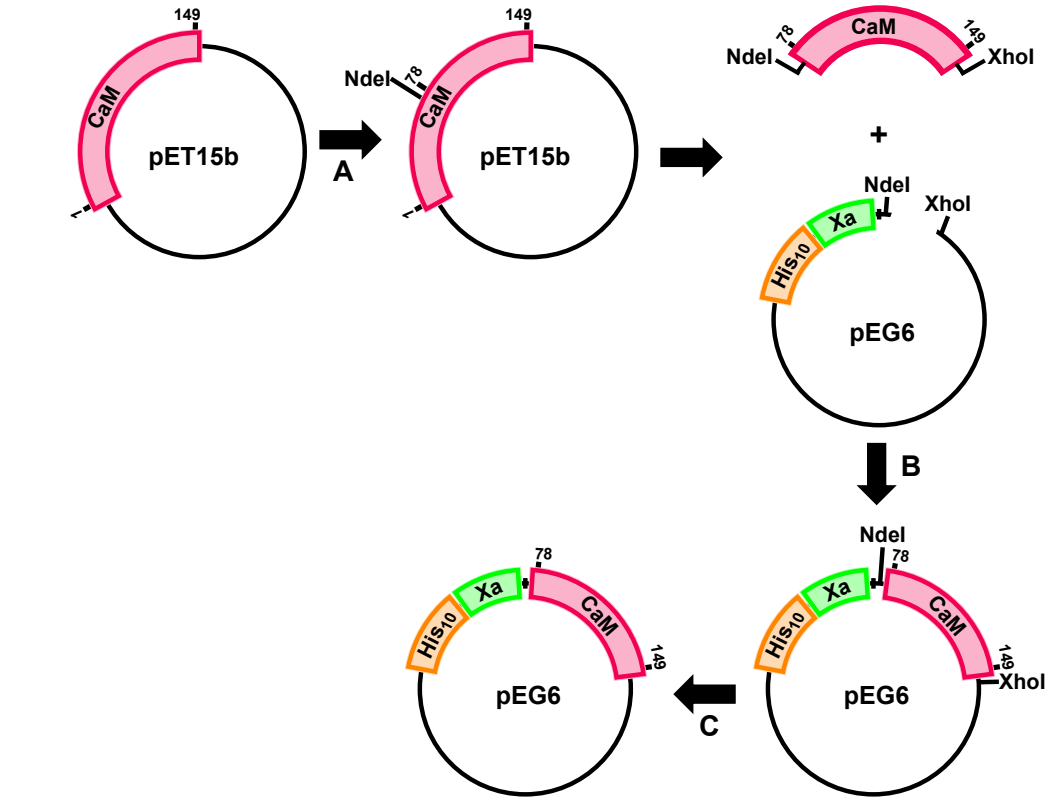
Figure 3-28. Construction of $\alpha S_{6\text{-}140}$ Plasmid

3.3.11 – Cloning of Calmodulin Expression Constructs

Quikchange ® mutagenesis was used to generate the following mutant plasmids: pET15b-H_{10}-$CaM_{77\text{-}149}$ (C-terminal domain) and pTXB1-$CaM_{1\text{-}72}$ (N-terminal domain). All mutagenesis steps were confirmed by DNA sequencing using the T7 promoter primer.

The C-terminal domain of calmodulin (CaM) was cloned from a plasmid containing a gene encoding full-length wild-type calmodulin from the chicken genus. This plasmid was provided by the Wand Laboratory of the University of Pennsylvania School of Medicine. The CaM gene had been cloned into the pET15b vector (Novagen, Gibbstown, NJ) at the NcoI and XhoI restriction enzyme cut sites. An NdeI site was

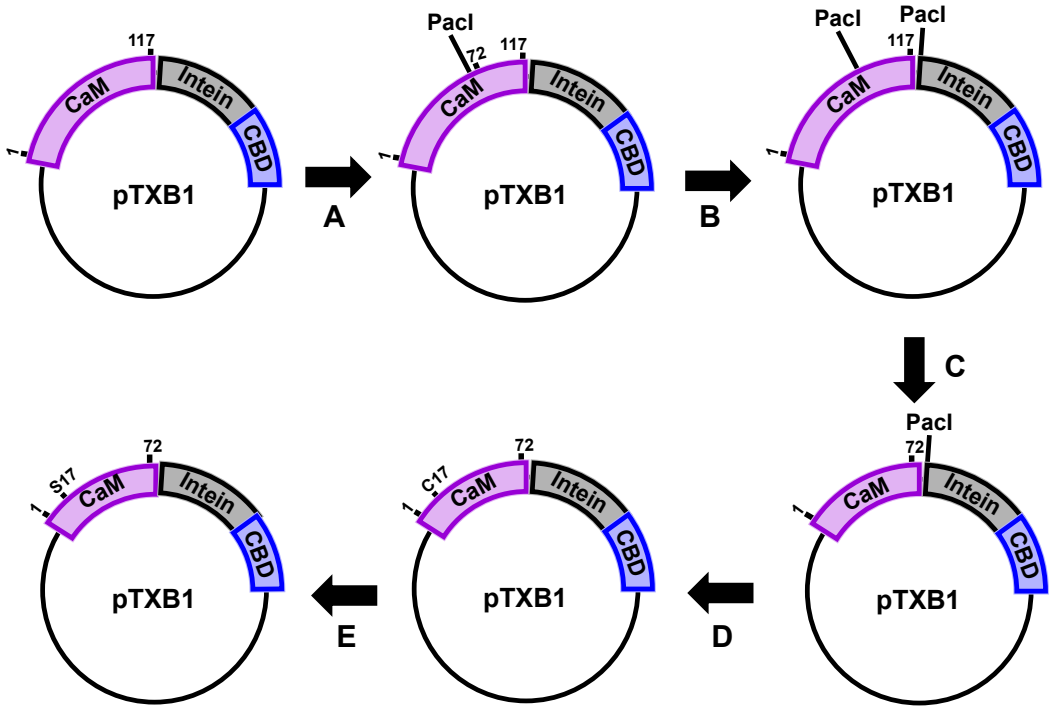
introduced at K_{76} of the CaM gene using QuikChange® mutagenesis. The plasmid was digested with PacI/NdeI. The CaM digested insert was ligated into a pEG6 vector[50], which contains an N-terminal H_{10} tag with a factor Xa cleavage site, at cut sites, Nde1 and XhoI with T4 DNA ligase. The pEG6 plasmid was a gift from the Muir Lab at Princeton University. To complete the plasmid, an N-terminal lysine was exposed through the deletion of HM at the inserted NdeI site resulting in the plasmid pEG6-H_{10}-FactorXa-(Lys)$CaM_{78\text{-}149}$ (Figure 3-29).



(A) Insert NdeI at K76	NdeI 5' – GACAATGATGGCAAGA**CATATG**AAAGATACAGATAGCG – 3'
(B) Create $CaM_{Nde1(76)\text{-}149}$ in pEG6	**Digest with NdeI/XhoI and ligate into pEG6 NdeI/XhoI vector**
(C) Delete HM at NdeI site	Mutant 5' – GCAGCGGCCATATCGAAGGTCGT AAAGATACAGATAGCGAAGAAG – 3' 5' – GCAGCGGCCATATCGAAGGTCGT**CATATG**AAAGATACAGATAGCGAAGAAG – 3' Native

Figure 3-29. Construction of WT $CaM_{78\text{-}149}$ His-Tagged Expression Plasmid

The N-terminal domain of calmodulin (CaM) was cloned from a plasmid containing a gene encoding wild-type calmodulin$_{S17C, 1-116}$. This plasmid was provided by the Linse Laboratory of the Department of Biochemistry and Structural Biology at Lund University. The plasmid contained the pTXB1 vector, which encodes a mini-intein GyrA from the *Mycobacterium xenopi* followed by a chitin-binding domain (CBD). A PacI site was introduced at M_{72} and L_{117} of the CaM gene using QuikChange® mutagenesis. With the PacI sites encoded, the plasmid was digested with PacI and then the resulting sticky ends were ligated using T4 DNA ligase resulting in a pTXB1-CaM$_{1-72}$-PacI plasmid. Next, the PacI site at 72 was mutated back to wild type M_{72}. To bring the CaM gene back to wild-type CaM$_{1-72}$, the S17C mutation was mutated back to wild type C17S (Figure 3-30).



(A) Insert PacI at M72	5' – CAATTGACTTCCCAGAGTTTC**TTAATTAA**ATGGCGCGCAAAATGAAAGATAC – 3'
(B) Insert PacI at L117	5' – GTGATGACAAACCTTGGTG**TTAATTAA**TGCATCACGGGAGATGCACTAG – 3'
(C) Create CaM_{1-72}	**Digest with PacI, remove CaM_{73-117}, ligate between PacI sites M72 and Intein**
(D) Mutate $PacI_{72}$ to WT (TM_{71-72})	5' – CAATTGACTTCCCAGAGTTTC**TGACAATG**TGCATCACGGGAGATGCAC – 3'
(E) Mutate C17S	5' – CAGAGTTCAAAGAGGCTTTT**AGC**CTG – 3'

Figure 3-30. Construction of WT CaM_{1-72} Intein Expression Plasmid with C-Terminal Chitin-Binding Domain

3.3.12 – Expression and Purification of αS_{6-140}

The plasmid containing the αS gene was used to transform *Escherichia coli* BL21(DE3) cells. Transformed cells were selected on the basis of ampicillin resistance. Single colonies were used to inoculate 5 mL of LB media supplemented with ampicillin (100 µg/mL). The primary 5 mL culture was incubated at 37 °C with shaking at 250 rpm

for 5 hours. 500 mL of LB media supplemented with ampicillin was inoculated with 1 mL of the primary culture. The 500 mL culture was incubated at 37 °C with shaking at 250 rpm until an OD_{600} of 0.6 was reached. Expression was induced with the addition of 1.0 mM isopropyl β-D-thiogalactoside and cells were incubated at 25 °C for 16 hours. The cells were harvested at 5000 x g for 15 min and the resulting pellet was resuspended in a Factor Xa cleavage buffer (100 mM NaCl, 50 mM Tris pH 8.0, 5 mM $CaCl_2$) and included protease inhibitor cocktail, 1 mM PMSF, and 10 units/mL DNAse1–Grade II) and sonicated. Following sonication, the cell lysate was boiled for 20 minutes prior to centrifugation for 20 minutes at 30,000 x g, 4 °C. Collected soluble protein was gently shaken for 1 h on ice with Ni-NTA resin. Protein was purified by rinsing with Factor Xa cleavage buffer followed by washing with Factor Xa wash buffer (100 mM NaCl, 50 mM Tris pH 8.0, 5 mM $CaCl_2$, and 50 mM imidazole). Protein was eluted with Factor Xa elution buffer 100 mM NaCl, 50 mM Tris pH 8.0, 5 mM $CaCl_2$,and 250 mM imidazole). SDS-PAGE analysis was performed to analyze the purity of αS elution fractions. Collected protein fractions were dialyzed against Factor Xa cleavage buffer at 4 °C. The His_{10} purification tag was removed by incubation of purified αS (0.5 mg/mL) with Factor Xa, Restriction Grade, Bovine Plasma (Novagen) (20 Units) overnight at 37 °C. Factor Xa was removed by boiling for 20 minutes and centrifugation for 20 minutes at 30,000 x g, 4 °C. Soluble protein was purified from N-terminal peptide and uncleaved protein by 1 h incubation with Ni-NTA resin. Protein concentration was determined by BCA assay (Pierce) and adjusted to 1.0 mg/mL, purity was confirmed by SDS-PAGE and MALDI mass spectrometry. Protein was stored at – 80 °C until further use.

3.3.13 – Expression and Purification of CaM_{78-149}

The plasmid containing the CaM_{78-149} gene was used to transform *Escherichia coli* BL21(DE3) cells. Transformed cells were selected on the basis of ampicillin resistance. Single colonies were used to inoculate 5 mL of LB media supplemented with ampicillin (100 μg/mL). The primary 5 mL culture was incubated at 37 °C with shaking at 250 rpm for ~ 3 hours, until log phase was reached. One liter of LB media supplemented with ampicillin was inoculated with 1 mL of the primary culture. The 1 liter culture was incubated at 37 °C with shaking at 250 rpm until an OD_{600} of 0.6 was reached. Expression was induced with the addition of 1.0 mM isopropyl β-D-thiogalactoside and cells were incubated at 25 °C for 16 hours. The cells were harvested at 5000 x g for 15 min and the resulting pellet was resuspended in binding buffer (50 mM Tris pH 8) and included protease inhibitor cocktail, 1 mM PMSF, and 10 units/mL DNAse1–Grade II) and sonicated. Following resuspension, the cell lysate was sonicated for 1 min, 6 times, with a one minute rest between each sonication. Collected soluble protein was gently shaken for 1 h on ice with Ni-NTA resin. Protein was purified by rinsing with Factor Xa cleavage buffer followed by washing with wash buffer (50 mM Tris pH 8 and increasing imidazole from 10 mM – 50 mM). Protein was eluted with elution buffer (50 mM Tris pH 8 and 250 mM imidazole). SDS-PAGE analysis was performed to analyze the purity of CaM_{78-149} elution fractions. Collected protein fractions were dialyzed against Factor Xa cleavage buffer (50 mM Tris, 100 mM NaCl, 5 mM $CaCl_2$ pH 8) at 4 °C. The His_{10} purification tag was removed by incubation of purified CaM_{77-149} with Factor Xa, Restriction Grade, Bovine Plasma (Novagen) (20 Units) overnight at 37 °C (or until reaction is complete, which can take up to 2 days if

high concentration of protein is used). Factor Xa cleavage was halted by addition of 1 mM PMSF, and solution was kept at room temperature for 1 h to ensure PMSF degradation. Protein concentration was determined by BCA assay (Pierce) and adjusted to 1.0 mg/mL, purity was confirmed by SDS-PAGE and MALDI mass spectrometry. Protein was stored at – 80 °C until further use.

3.3.14 – Expression and Purification of CaM_{1-72}

The plasmid containing the CaM_{1-72} gene was used to transform *Escherichia coli* BL21(DE3) cells. Transformed cells were selected on the basis of ampicillin resistance. Single colonies were used to inoculate 5 mL of LB media supplemented with ampicillin (100 μg/mL). The primary 5 mL culture was incubated at 37 °C with shaking at 250 rpm for ~ 3 hours, until log phase was reached. One liter of LB media supplemented with ampicillin was inoculated with 1 mL of the primary culture. The 1 liter culture was incubated at 37 °C with shaking at 250 rpm until an OD_{600} of 0.6 was reached. Expression was induced with the addition of 1.0 mM isopropyl β -D-thiogalactoside and cells were incubated at 25 °C for 16 hours. The cells were harvested at 5000 x g for 15 min and the resulting pellet was resuspended in binding buffer (20 mM Tris, 0.5 M NaCl pH 7.5) and included protease inhibitor cocktail, 1 mM PMSF, and 10 units/mL DNAse1–Grade II) and sonicated. Following resuspension, the cell lysate was sonicated for 1 min, 6 times, with a one-minute rest between each sonication. Collected soluble protein was incubated with Chitin Beads (pre-equilibrated with binding buffer) and gently rotated at 4 °C for 1 - 4 hours (longer incubation enhanced chitin binding to protein). Post incubation, the flow through was collected and the column was washed

with 3 x 10 mL binding buffer. 2.5 mL of chitin cleavage buffer was added to the beads (20 mM Tris, 0.5 M NaCl, 200 mM MESNA pH 7.5) and incubated at 4 °C with rotation. Protein was eluted with elution buffer by collecting incubated flow through and beads were washed with 3 x 2 mL chitin cleavage buffer. SDS-PAGE analysis was performed to analyze the purity of CaM_{1-72} elution fractions. Collected protein fractions were dialyzed against water overnight at 4 °C. Protein concentration was determined by BCA assay (Pierce) and adjusted to 1.0 mg/mL, purity was confirmed by SDS-PAGE and MALDI mass spectrometry. Protein was stored at – 80 °C until further use.

3.3.15 – Full Enzymatic Transfer to αS_{6-140} and CaM_{78-149}

For a typical reaction, 4.25 mg αS_{6-140} in AaT Ligation Buffer pH 8.0 was modified with Hcm (1 mM) containing 20 mg *E. coli* total tRNA, 2.5 mM ATP, 1.0 - 0.1 mg *E. coli* Met*RS, and 0.1 mg *E. coli* AaT. The reaction was carried out at 37 °C for 4 h and arrested by boiling for 30 min. The supernatant was cleared by centrifuging for 20 min at 30,000 x g. Hcm-αS_{6-140} (**19**) was further purified on the with a HiTrap Q HP column (GE Healthcare). Protein was buffer exchanged into Milli-Q water and stored at -80 °C. The same procedure was used for CaM_{78-149} except that post-reaction the sample was not boiled and instead the reaction was halted by dialysis against 20 mM Tris pH8 overnight at 4 °C. FPLC purification and purified protein storage remained the same for CaM as with αS.

3.3.16 – Ligation of Ac-MetAspValPhe-SR (11) and Ac-McmMetAspValPhe-SR (S10) to Hcs-αS$_{6-140}$ (19)

Ligation buffer was prepared fresh as follows: 100 mM sodium phosphate dibasic, 20 mM TCEP, 6 M guanidine hydrochloride and 2 % v/v thiophenol at pH 7.2. The buffer was sparged twice with argon: 20 minutes for a 100 mM sodium phosphate dibasic and 6 M guanidine hydrochloride solution and then for an additional 10 min post addition of TCEP, thiophenol, and pH adjustment to 7.2. Sparged ligation buffer (100 uL) was added to 0.0404 μmol (1 equivalent) of dried Hcm-αS$_{6-140}$ protein. The mixture was briefly vortexed and allowed to reduce for 10 min. Next, the protein solution was added to 0.1212 μmol (3 equivalents) of Ac-MetAspValPhe-SR (**11**). The reaction was mixed overnight at 250 RPM at 37°C. The reaction was quenched the following day with a 3 mL addition of MilliQ water and buffer exchanged solvent with six water washes using a 3 kDa protein concentrator microfuge tube and analyzed by HPLC (Gradient 7). Protein retention times were monitored at 215 nm (and 325 nm for **S13**). MALDI MS analyses are in Table 3-2. Example HPLC traces are shown in Figure 3-8.

3.3.17 – Ellman's Reagent Analysis

0.8 nmoles of αS-Mcm-Hcs$_5$ (**20a**) were methylated with 1000 equivalents of methyl iodide in 0.1 M sodium bicarbonate pH 8.6 for 5 min at room temperature. Following methylation, the reaction was incubated with 150 μM 5,5'-dithiobis-(2-nitrobenzoic acid (DTNB) for 5 min at room temperature. UV/Vis measurements were taken using a Tecan M1000 plate reader. Excitation slit width was 5 nm, integration time 0.02 s, data interval 2.0 nm, flash frequency 400 Hz, 25 flashes, gain set to 100, and z-

height 20 mm (Figure 3-7). Section 3.3.17 was performed by John Warner.

3.3.18 - aS-Mcm Fluorescence Readings

Fluorescence spectra were collected with a Varian Cary Eclipse fluorescence spectrophotometer fitted with a Peltier multicell holder. Fluorescence spectra were collected using quartz cells with 1 cm path lengths. Excitation and emission slit widths were 5 nm, scan rate 120 nm/min, averaging time 0.5 s, and data interval 1.0 nm. Section 3.3.18 was performed by John Warner.

3.3.19 – Trypsin Digest and MALDI MS Analysis

Ac-αS **21b** (5.4 μg) was buffer-exchanged into 100 μL of freshly prepared 25 mM NH_4HCO_3 (pH 7.6), and digested with 2 μL of sequencing-grade modified trypsin (0.1 μg/μL). The digestion was allowed to proceed at 37 °C for 3 h. An aliquot (1.0 μL) of the digest was taken and analyzed by MALDI MS in CHCA matrix (Figure 3-31, Table 3-4). MALDI MS data shown in Figure 3-32 confirm that no methylation of His_{50} took place. To determine the sensitivity of our method for identifying the extent of Hcs methylation, we synthesized authentic Ac-MetAspValPheHcsLys (**S8**) and Ac-MetAspValPheMetLys (**S9**). We mixed them in ratios varying from 1:1 to 1:100 **S8/S9** and spotted these mixtures in CHCA matrix in a manner identical to the analysis of trypsinized **21b**. We found that the ionization of Hcs-containing **S8** was proportional to its ratio in the mixture. Therefore we feel confident that ratios of Ac-MetAspValPheHcsLys to Ac-MetAspValPheMetLys observed in the MALDI MS data of trypsinized **21b** are reflective of the degree of conversion of Hcs_5 to Met_5. The 5:95 and

1:99 **S8/S9** MALDI MS data are shown in Figure 3-31.

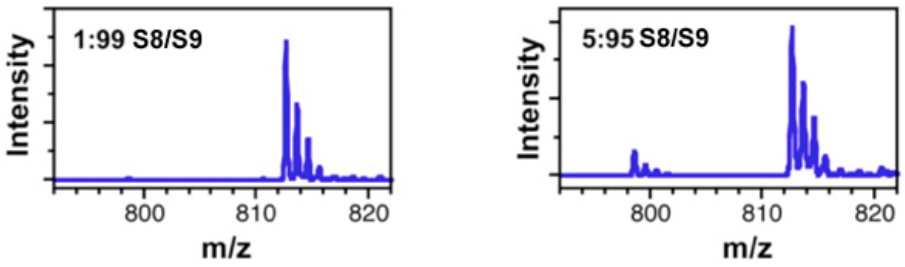


Figure 3-31. MALDI MS Data for Mixtures of Ac-MetAspValPheHcsLys (S8) and Ac-MetAspValPheMetLys (S9)

Table 3-4. Trypsin Digest Ac-αS (13b) Sequence Coverage

m/z	Start	End	Missed	Peptide
798.42	1	6	0	Ac-MDVFXK
812.37	1	6	0	Ac-MDVFMK
830.44	24	32	0	QGVAEAAGK
873.47	13	21	0	EGVVAAAEK
951.51	35	43	0	EGVLYVGSK
1295.70	46	58	0	EGVVHGVATVAEK
1478.78	81	96	0	TVEGAGSIAAATGFVK
2157.19	59	80	1	TKEQVTNVGGAVVTGVTAVAQK
4286.73	103	140	0	NEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEPEA

Underline indicates observed fragment. Ligation site fragment shown in bold. X = Homocysteine (Hcs).

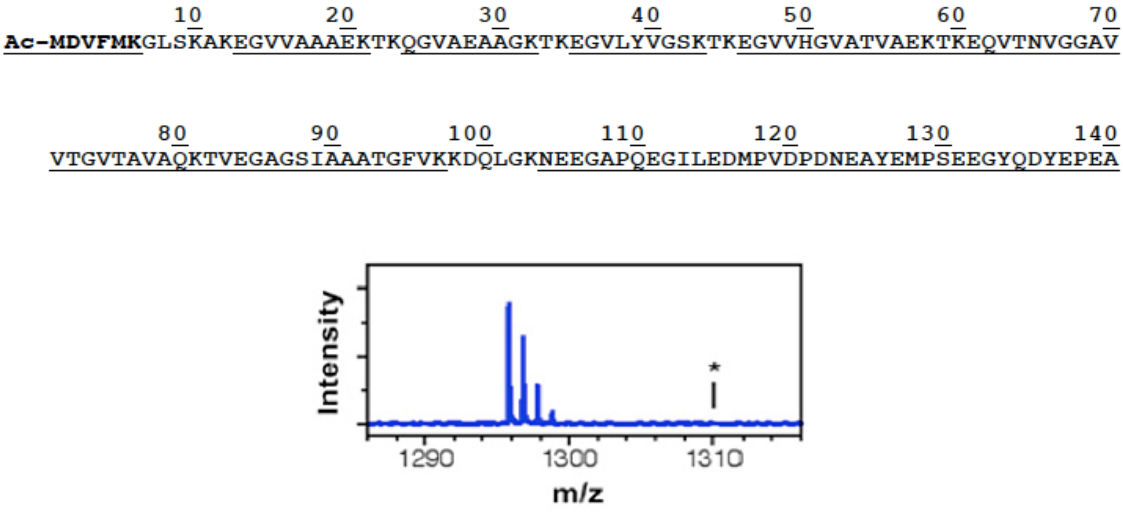


Figure 3-32. αS Full-Length Sequence and MALDI MS Data for Trypsin Digested Ac-αS$_{6-140}$
The fragment corresponds to EGVVHGVATVAEK and the asterisk indicates mass of peptide with methylated His$_{50}$.

3.3.20 – Ligation Motif Prevalence Analysis

The frequency of XaaMetArg and XaaMetLys motifs (Xaa = any amino acid) in protein sequences of the Protein Data Bank (PDB) was assessed using PHI-BLAST.[174] Briefly, a four-letter query, "X_1MRX_2" or "X_1MKX_2" (where X_1 and X_2 were varied to cover all 20 natural amino acids), was aligned to the database under the constraints of a "X_1MR/K" pattern and the PAM30 substitution matrix. The results were then pooled in a spreadsheet, inspected for false hits, and categorized (Table 3-5, 3-6 and 3-7) by Jerri Wang.

Table 3-5. Summary of Unique PDB Entries Containing One or More MR or MK Ligation Site(s)

Ligation Sites per Entry	# of Entries	Ligation Sites per Entry	# of Entries	Ligation Sites per Entry	# of Entries
1	13,550	**5**	142	**9**	5
2	4,842	**6**	44	**11**	1
3	1,554 7	**14**	14	**1**	1
4	543	**8**	4	**17**	1
of unique entries with MetArg or MetLys Ligation site(s)					**20,701**
# of unique entries with N-Terminal (≤ 40 aa) MetArg or MetLys Ligation site(s)					**5,156**
# of entries in the PDB Database (as of Oct 14, 2012)					**60,325**

Table 3-6. Selected Examples of Potential Ligation Targets

Protein	Taxonomy	Length Alignment	PDB ID
Human Prion Protein	*H. sapien*	142 19 NMK 21	2LSB
Ribonuclease S	*H. sapien*	127 6 MMK 8	2K11
Caspase 2, Chain B	*H. sapien*	105 29 AMR 31	1PYO
CTerminal Src Kinase	*H. sapien*	278 19 NMK 21	1BYG
Annexin	*H. sapien*	319 26 AMK 28	2ZOC
Anti HIV1 Protease Fab, Chain H	*M. musculus*	226 29 FMR 31	1MF2
Cytochrome P450	*B. megaterium*	471 29 LMK 31	1FAH
Myosin	*P. magellanicus*	840 20 MMK 22	2OS8
BetaAlanine Synthase	*S. kluyveri*	474 34 GMR 36	2VL1
30S Ribosomal Protein S2	*E. coli*	241 26 KMK 28	1VS5

Table 3-7. Distribution of Ligation Sites by Thioester Residues and Relative Positions

Thioester Residue	# of Distinct Ligation Sites		Relative Positions					
			N-Terminal (≤ 40 aa)		Middle		C-Terminal (≤ 40 aa)	
	MR	MK	MR	MK	MR	MK	MR	MK
Very Suitable for Native Chemical Ligation								
G	846	1,158	102	202	644	815	100	141
A	1,224	1,808	139	354	961	1,177	124	277
L	1,261	1,691	135	234	945	1,183	181	274
M	411	524	72	169	242	268	97	87
F	542	654	78	106	375	455	89	93
Y	506	590	38	74	401	427	67	89
Sub-total	**11,215**		**1,703**		**7,893**		**1,619**	
Suitable for Native Chemical Ligation								
K	729	1,109	138	170	481	754	110	185
R	828	898	125	181	589	579	114	138
N	529	619	83	122	389	410	57	87
Q	528	692	115	105	354	509	59	78
S	932	1,054	266	327	579	551	87	176
C	202	224	28	30	129	156	45	38
W	189	197	23	18	139	166	27	13
Sub-total	**8,730**		**1,731**		**5,785**		**1,214**	
Mutations Needed to Promote Native Chemical Ligation								
V	1,029	1,155	164	141	706	831	159	183
I	702	941	80	147	488	684	134	110
T	784	950	124	150	516	647	144	153
H	371	761	211	348	137	329	23	84
D	636	905	59	104	442	663	135	138
E	907	1,391	131	195	601	952	175	244
P	396	386	52	65	278	257	66	64
Sub-total	**11,314**		**1,971**		**7,531**		**1,812**	
Total	**31,259**		**5,405**		**21,209**		**4,645**	

CHAPTER 4

EXPANSION OF AMINOACYL TRANSFERASE SUBSTRATE SCOPE

Section 4.1 - Introduction

An enzyme's active site is the location in which a chemical reaction occurs, and the amino acid residues within the active site convey binding to appropriate substrates. Many factors dictate whether a compound will be an enzyme substrate, including substrate/enzyme complementarity, active site flexibility, steric matching, and noncovalent forces arising from hydrophobicity and charge.[131] Some active sites have stringent substrate restrictions, which have evolved to make the enzyme highly specific to one substrate. Alternatively, an enzyme can have less specificity, which would allow it to accept a variety of substrates; *E. coli* leucyl/phenylalanyl aminoacyl transferase (AaT) falls into this second class. AaT is an enzyme that can naturally transfer three amino acids, leucine, phenylalanine and – to a lesser extent – methionine, from their respective aminoacyl tRNAs, to the α-amino group of proteins with an N-terminal lysine or arginine.[126] The active site of AaT is very hydrophobic and also selective against β-branched amino acids, such as valine.[56] The enzyme delicately balances moderate selectivity for its amino acid substrate with high specificity binding for the protein N-terminus.

Using X-ray crystallography, MALDI-TOF MS, amino acid mutagenesis, and enzyme kinetics, two groups have proposed mechanisms for the AaT amide bond-forming reaction. Suto *et al*. solved a crystal structure of AaT bound to an aminoacyl-tRNA analog, puromycin, to help elucidate the key residues that recognize the aminoacyl-tRNA substrate. They performed site-directed mutagenesis at residues within close proximity to puromycin bound in the enzyme active site. A hydrophobic pocket

containing, M_{144}, F_{153}, L_{170}, F_{173}, and I_{185}, was found to hold the *p*-methoxybenzyl moiety of puromycin, which represented the phenylalanine/leucine portion of the natural aminoacyl substrate (Figure 4-1).[56] Upon mutation of those residues to alanine, they found enzymatic activity to be substantially reduced, indicating the importance of the specific size of the hydrophobic amino acids in the active site to correctly fit the Phe/Leu aminoacyl tRNA substrates. Additionally, the hydrophobic residues are well conserved within transferases from other eubacteria, giving further proof of their importance in the AaT amide bond-forming reaction.[56] Additional residues that were proposed to be linked to AaT activity included Glu_{156} and Gln_{188}, which could interact with the positively-charged N-terminal arginine/lysine of the protein substrate by positioning it for favorable reactivity.[43]

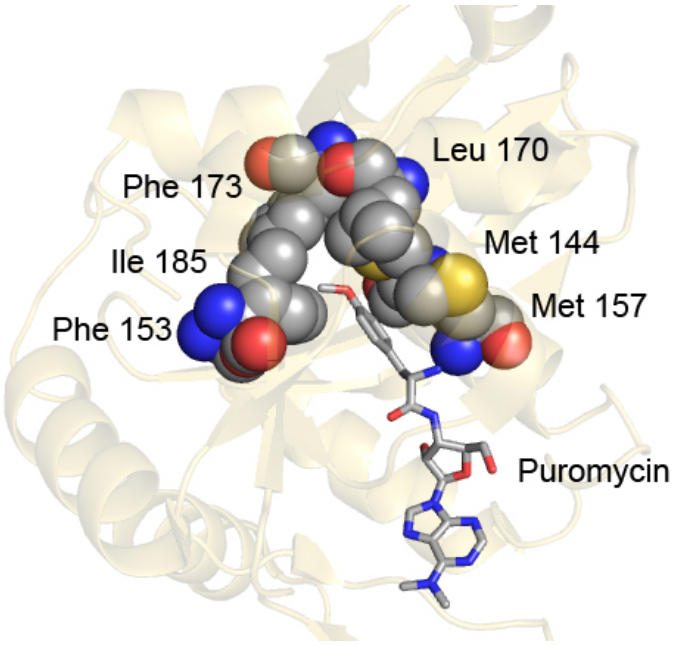


Figure 4-1. Recognition of Puromycin in AaT
From X-ray crystal structure by Suto *et al*. Hydrophobic pocket: Met_{144}, Phe_{153}, Leu_{170}, Phe_{173}, Ile_{185}.[56]

The same group proposed a specific AaT reaction mechanism in a following paper in 2007 when they obtained two crystal structures of AaT, one with phenylalanyl adenosine, the 3' end of the phenylalanine aminoacyl tRNA natural substrate, and one with the product Phe-ligated peptide (PheArgTyrLeuGly).[43] Watanabe *et al*. proposed that the AaT ligation reaction of phenylalanine from Phe-tRNA to the N-terminus of an Arg/Lys-peptide had a catalytic mechanism similar to the reverse reaction of a serine protease (Figure 4-2).[43] They determined that nucleophilic attack of the backbone α-amino group of the N-terminal arginine/lysine on the carbonyl carbon of the 3' ester bond of the aminoacyl tRNA is facilitated by several amino acid residues within the active site. They proposed that Gln_{188} acts as a general base, while Asp_{186} helps place Gln_{188} in the correct reaction orientation, and Asn_{191} stabilizes the formation of the oxyanion hole that is produced in the formation of the tetrahedral intermediate. When comparing the product-bound peptide structure to the starting material Phe-A bound state, they found substantial amino acid movement had occurred within the active site, where tryptophan residues W_{49} and W_{111} rotated and flipped, respectively, to accommodate the product. As they had found before, in addition to the hydrophobic pocket that holds the phenylalanine portion of the aminoacyl-tRNA substrate, the specific size of the negatively charged site that holds the arginine in place is directly connected to AaT activity, with alanine or aspartate mutations abolishing or reducing activity, respectively. Their research determined that the AaT reaction does not act like the peptidyl transfer site of the ribosome, which positions residues for substrate-assisted catalysis (below), and instead mimics a reverse acylation reaction seen in proteases and esterases.

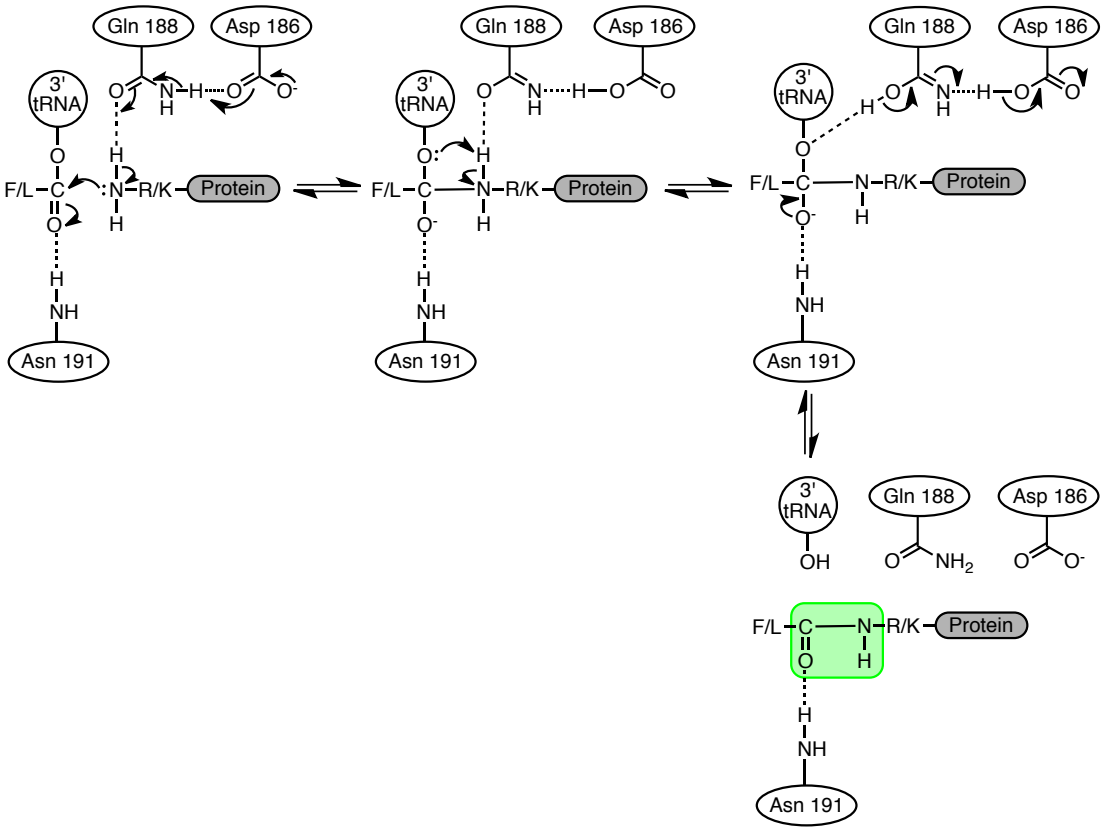


Figure 4-2. First Proposed AaT Ligation Mechanism
Adapted from Watanabe *et al*. 2007.[43]

The mechanism proposed by Watanabe e*t al*. was challenged by Fung *et al*. in 2011.[59] They reevaluated the catalytic mechanism using a quantitative MALDI-TOF MS-based assay with radioisotope-enriched peptide standards. Fung *et al*. stated that it was unconventional for glutamine (Q_{188}) to act as general base and therefore re-analyzed the alanine mutations with their MALDI method to determine if the residue mutants were indeed catalytically deleterious.[59] Using their highly sensitive technique, they determined that the previously-determined catalytic residues, D_{186} and Q_{188}, were not directly involved with catalysis, but instead critical for correct orientation of the aminoacyl tRNA

and acceptor protein, which react to form a peptide bond without the direct aid of the enzyme. This is similar to the established mechanism of peptide bond formation on the ribosome (Figure 4-3).[59] An additional argument that they make for a ribosome-like active site behavior is the slow reaction rate of AaT, $k_{cat}/K_M = 8.02 \times 10^3\ M^{-1}s^{-1}$.[43, 59] The substrate-assisted catalysis mechanism is still a proposal, however, and the exact influence of AaT on peptide bond catalysis is currently unconfirmed. While the two proposed mechanisms suggest differing means of enzyme catalysis, both include nucleophilic attack on the aminoacyl tRNA carbonyl by the α-amino group arginine.

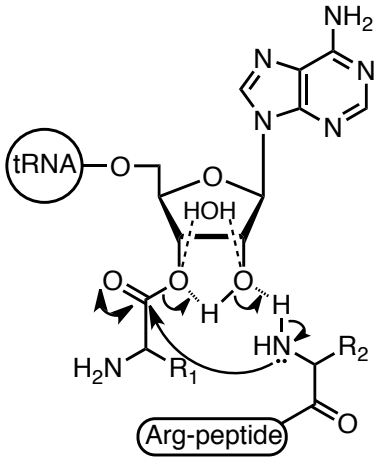


Figure 4-3. Second Proposed AaT Ligation Mechanism
Adapted from Fung *et al*. 2011. The reaction takes place *via* the 2'-hydroxyl in "substrate-assisted catalysis." The AaT-bound aminoacyl tRNA acts similarly to the P-site in the ribosome, and the Arg-peptide is analogous to the A-site.[59]

The AaT crystal structures with bound substrate Phe-A and product Phe-peptide, give evidence that the AaT active site has flexibility upon substrate binding; however, we had determined that the WT AaT, while accommodating a variety of unnatural amino acid residues, has a steric limit. In the aforementioned discussion of the AaT catalysis mechanism, mutagenesis was used to help determine key residues involved in substrate

recognition and influence in enzyme activity. For our purposes, we used alanine mutagenesis to expose space within the AaT active site to accommodate larger amino acid residues. We synthesized several aminoacyl adenosine analogs with residues that could not transfer to the N-terminus of our model peptide, LysAlaAcm, using WT AaT (Figure 4-4). We determined that we would need space within the active site to allow for extension in areas that around the aminoacyl adenosine α-amino group and sidechain R-group. By inspection of the crystal structures, we chose several amino acids that could "open up" the active site to accommodate the larger unnatural amino acids: $M_{144}A$, $M_{158}A$, $L_{170}A$, $F_{173}A$, $F_{177}A$, $I_{185}A$, $C_{187}A$. We made individual and double amino acid substitution combinations but had difficulty with expressed protein stability. We obtained several expression-stable single alanine mutants, $M_{144}A$, $L_{170}A$, $I_{185}A$, and $C_{187}A$, which we assayed for enzymatic activity with a natural aminoacyl substrate mimic, Phe-A, and corresponding donors with the unnatural amino acids, Mef, Azf, Nap, Nbd, Mcm, 1, 8-naphthalimide (Npt1,8), and 2, 3-Naphthalenedicarboximide (Npt2,3) (Figure 4-4). Phe-A and Nap-A had lower peptide bond-forming efficiency than WT and none of the unnatural residues larger than napthylalanine were compatible with our mutant enzymes.

We next established a collaboration with the Saven group to attempt to find an AaT mutant that, through predictions from computer-aided protein design, would be more active for one particular type of unnatural aminoacyl adenosine substrate, homocysteine or cysteine disulfide-protected analogs. An improved enzyme for these specific substrates would enhance our efficiency in forming native chemical ligation-

ready precursor proteins. Along with our studies of the AaT active site, we explored the possibility of utilizing another aminoacyl transferase, Bpt from *V. vulnificus*, by examining its ability to accept a minimized substrate, leucyl adenosine, similarly to AaT.

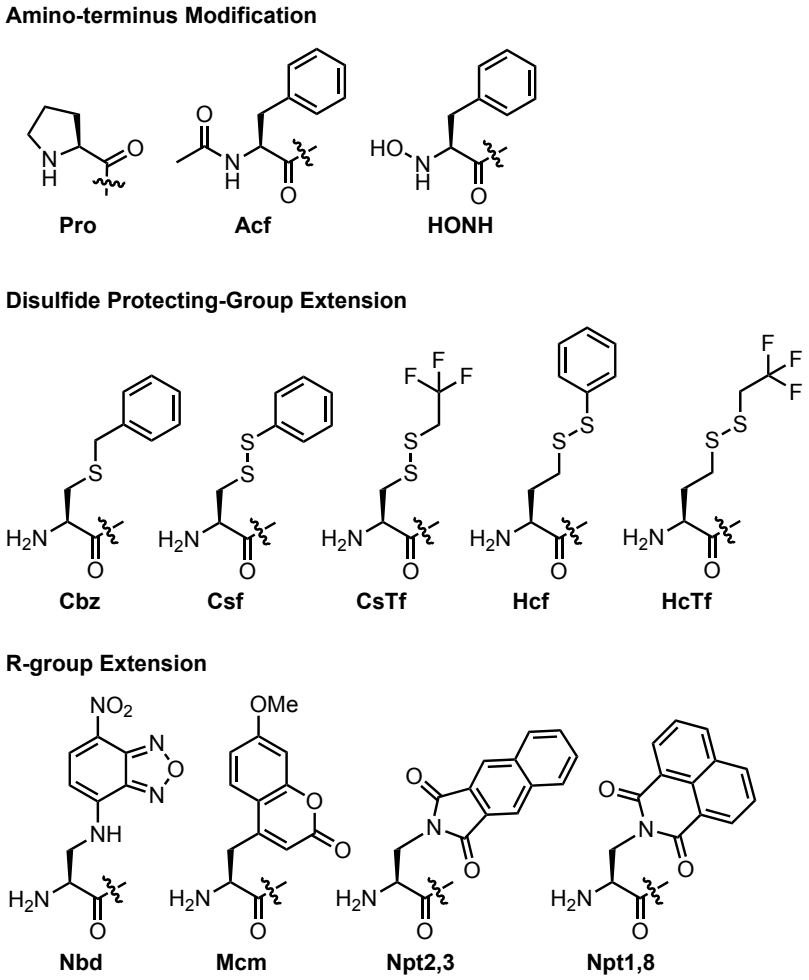


Figure 4-4. Amino Acids Not Utilized by WT AaT

Section 4.2 - Results and Discussion

In our previous work, we determined that WT AaT could use a variety of substrates with bulky sidechain R-groups such as naphthylalanine, and reactive handles residues such as *p*-azido-phenylalanine. However, we determined that the active site has

polar and steric restrictions; therefore, a phase of this project has been to create active site mutations to open space for more bulky and/or more polar R-groups in order to expand substrate scope. We chose a several amino acids that could "open up" the active site to accommodate the larger unnatural amino acids: $M_{144}A$, $M_{158}A$, $L_{170}A$, $F_{173}A$, $F_{177}A$, $I_{185}A$, $C_{187}A$ (Figure 4-5). We were careful not to choose amino acids that were directly involved with substrate binding in an attempt to avoid issues of enzyme catalysis inactivation. We made individual amino acid substitutions and additionally made all the respective combinations of double mutants of the listed residues, resulting in a total of 27 mutants. All of the residues chosen for alanine mutagenesis are located around the aminoacyl R-group recognition site. Phe_{173} and Phe_{177} are the farthest from the substrate phenylalanyl moiety, seen in Figure 4-5, but could possibly open a "slot" to fit planar fluorescent molecules such as Npt1,8 and Npt2,3.

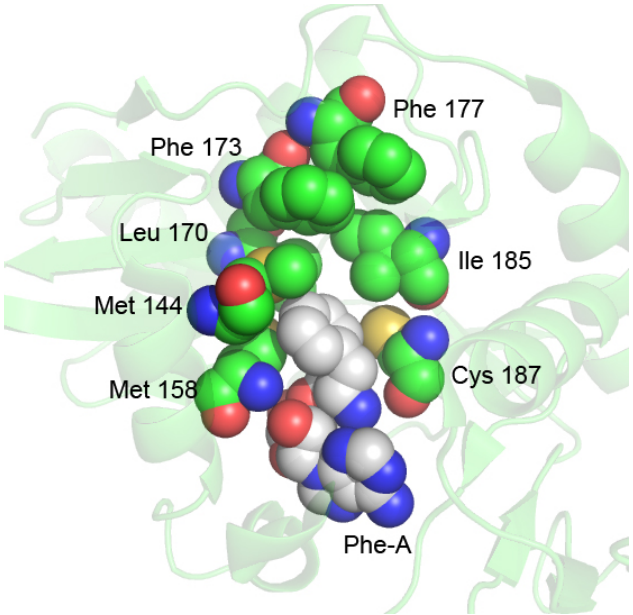


Figure 4-5. AaT Residues Selected for Mutagenesis
Residues to be mutagenized (green spheres) and Phe-A (grey spheres) bound to active site. PDB ID: 2Z3K[43]

Of the mutants, few could be expressed successfully. Most of the mutants had problems of folding stability, and while could be expressed, would fall out of solution upon cellular extraction mixed with the cellular debris and likely stuck in inclusion bodies (Figures 4-6 and 4-7). Additionally, with the mutants that expressed poorly, we tended to see an increase in a higher molecular weight band, which approximately corresponded to dimerization of AaT (Figure 4-7, beige arrow). Of the 27 mutant enzymes, enzymes that expressed to a degree comparable to wild type included both double mutants and single mutants. One double mutant in particular, $L_{170}A$ $M_{158}A$, had the best expression of the mutants (Figures 4-8 and 4-9). However, even with this case, a substantial amount of protein precipitated from solution during lysis and removal of cellular debris through centrifugation to obtain the soluble protein in the supernatant.

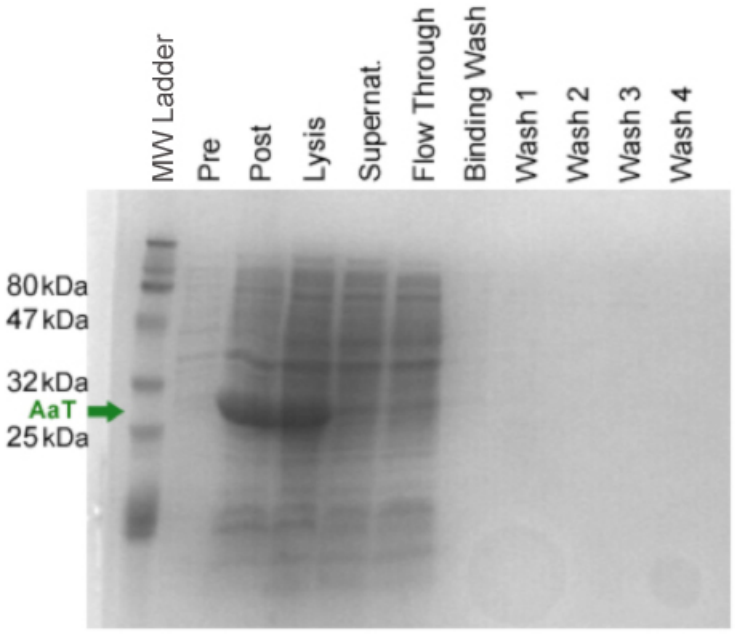


Figure 4-6. SDS PAGE Gel Production Analysis of $F_{177}A$ AaT

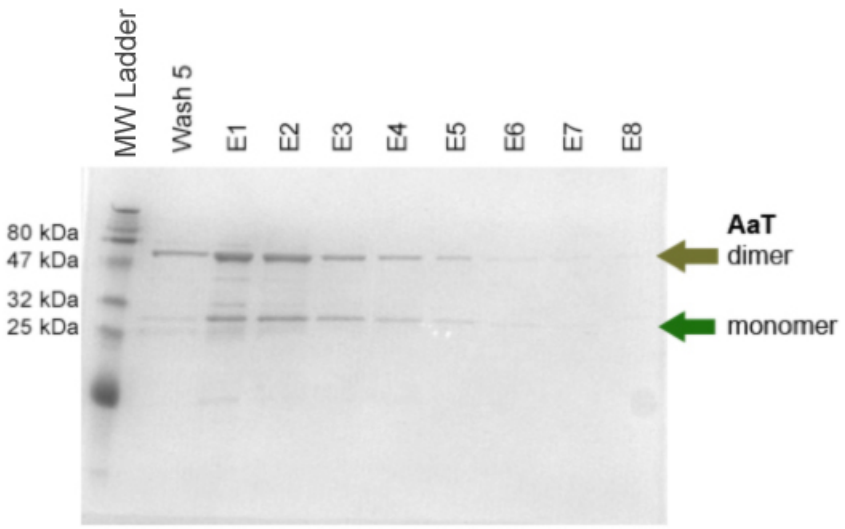


Figure 4-7. SDS PAGE Gel Expression Analysis of $F_{177}A$ AaT

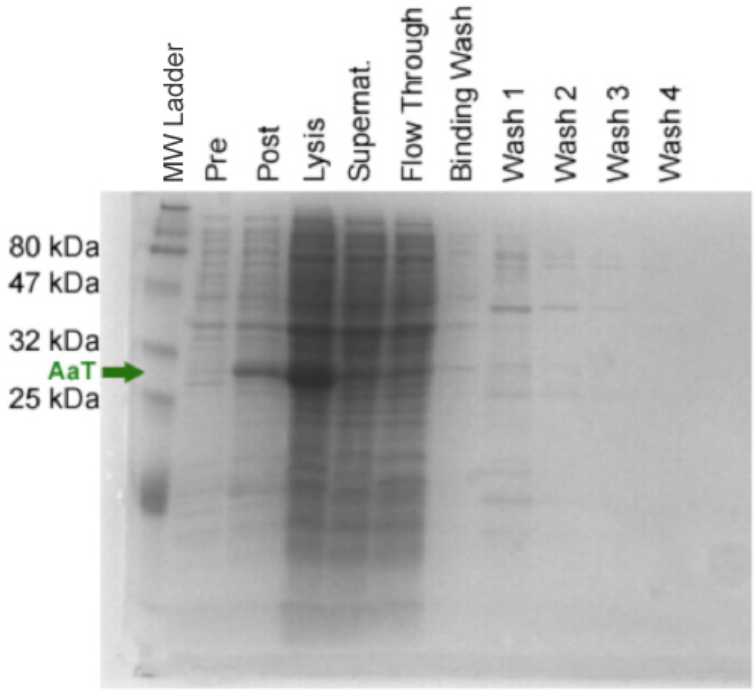


Figure 4-8. SDS PAGE Gel Production Analysis of $L_{170}A/M_{158}A$ AaT

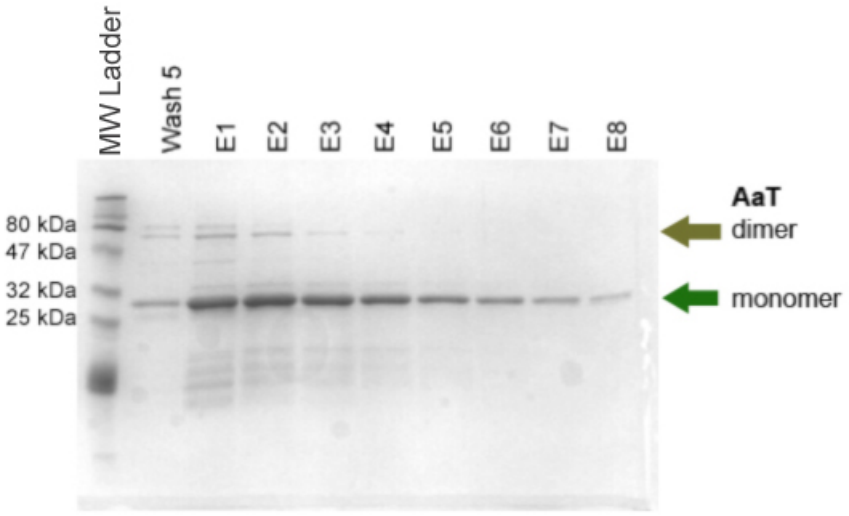


Figure 4-9. SDS PAGE Gel Expression Analysis of Double Mutant $L_{170}A/M_{158}A$ AaT

We attempted to keep the protein more soluble by altering lysis buffer conditions through changing the salt composition (Figure 4-10). We had noticed that we lost some protein into the insoluble fraction during overexpression; therefore, due to the expression variability of each mutant we did this lysis analysis on WT AaT. The standard resuspension lysis buffer conditions consist of 50 mM Tris, 10 mM imidazole, 300 mM potassium chloride, and 5 mM β-mercaptoethanol. For each altered condition, DNAse, PMSF, and His-Tag protease inhibitor concentrations were kept the same as for standard WT AaT overexpression. The conditions varied with salt concentration ranging from 300 – 600 mM of ammonium sulfate or potassium chloride. Additionally, one condition maintained standard lysis buffer conditions with the addition of 10 % glycerol. None of these changes increased the percentage of soluble AaT retained in the supernatant.

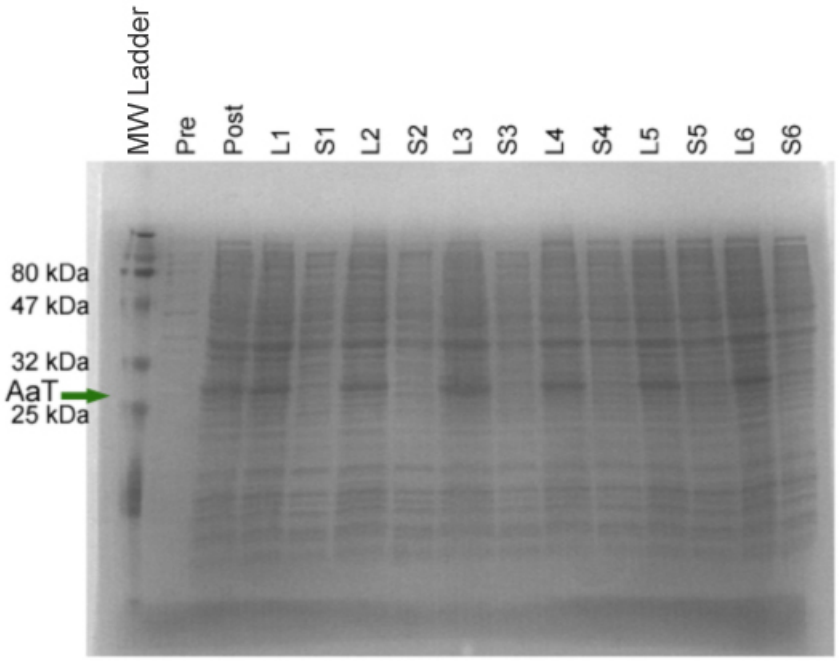


Figure 4-10. Soluble Fraction Analysis of Salt Concentration Lysis Conditions for WT AaT
L = lysis, S = Supernatant. Conditions (1) 50 mM Tris, 10 mM imidazole, 300 mM ammonium sulfate, 5 mM β-mercaptoethanol, (2) 50 mM Tris, 10 mM imidazole, 600 mM ammonium sulfate, 5 mM β mercaptoethanol, (3) 50 mM Tris, 10 mM imidazole, 600 mM potassium chloride, 5 mM β-mercaptoethanol, (4) 50 mM Tris, 10 mM imidazole, 300 mM potassium chloride, 300 mM ammonium sulfate, 5 mM β-mercaptoethanol, (5) 50 mM Tris, 10 mM imidazole, 300 mM potassium chloride, 5 mM β-mercaptoethanol, 10 % glycerol, (6) Standard resuspension lysis buffer conditions: 50 mM Tris, 10 mM imidazole, 300 mM potassium chloride, 5 mM β-mercaptoethanol.

After the first trial of lysis buffers, we changed other possible variables that could affect AaT solubility, including the duration and type of cell lysis, higher percentages of glycerol in the lysis buffer, and different buffer compositions used in published AaT overexpression protocols. We tested seven different conditions for cell lysis (Figure 4-11). The trials that yielded the greater soluble protein over wild type conditions included increased waiting time between 1 min sonications, addition of 30 % glycerol to the standard lysis buffer, or use of lysozyme instead of sonication to break open the cells. However, it is important to note that in all of the higher yielding AaT protein

supernatants, a substantial amount of AaT was still lost between the lysis and supernatant removal steps.

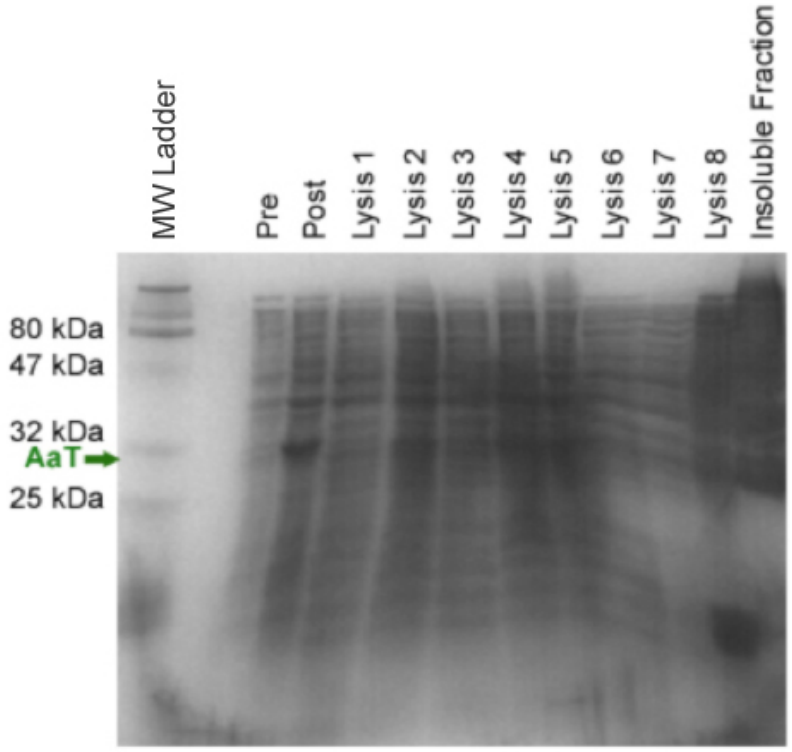


Figure 4-11. Soluble Fraction Analysis of Alternate AaT Lysis Conditions
Lysis buffer conditions are as follows: (1) decreased sonication to three iterations of 1 min sonication at 30 %, (2) Increased waiting time between standard six sonications from 1 min to 2 min, (3) included 20 % glycerol in lysis buffer, (4) included 30 % glycerol in lysis buffer, (5) new lysis buffer = 50 mM Tris, 1 mM imidazole, 100 mM KCl[130] (6) new lysis buffer = 20 mM Tris, 500 mM NaCl[175] (7) used lysozyme for cell lysis instead of sonication. Unless otherwise specified, all resuspended cells were lysed by sonication as described in 4.3 materials and methods section for WT AaT.

Given that the aforementioned buffer and lysis conditions did not substantially decrease the amount of AaT lost into the cell debris pellet, we decided to take another approach, and completely change the cell overexpression conditions. The overexpression protocol was adapted from Graciet *et al*.[50] The major changes of this protocol were a one

hour pre-IPTG cell growth stage at 42 °C and a post IPTG-induced cell growth phase for a shortened period, 5 h instead of 16 h (Figures 4-12 and 4-13). We found that our new overexpression protocol had better retention of soluble protein compared to previous mutant protein expressions. However, again, much of the overexpressed AaT was lost in the cell pellet again during cell debris removal. Therefore, despite our best efforts, we decided to continue our analysis of larger unnatural aminoacyl adenosine analogs substrates with AaT mutants that had previously expressed to a high enough concentration to be tested by our LysAlaAcm peptide bond formation analysis assay (see 4.3 Materials and Methods).

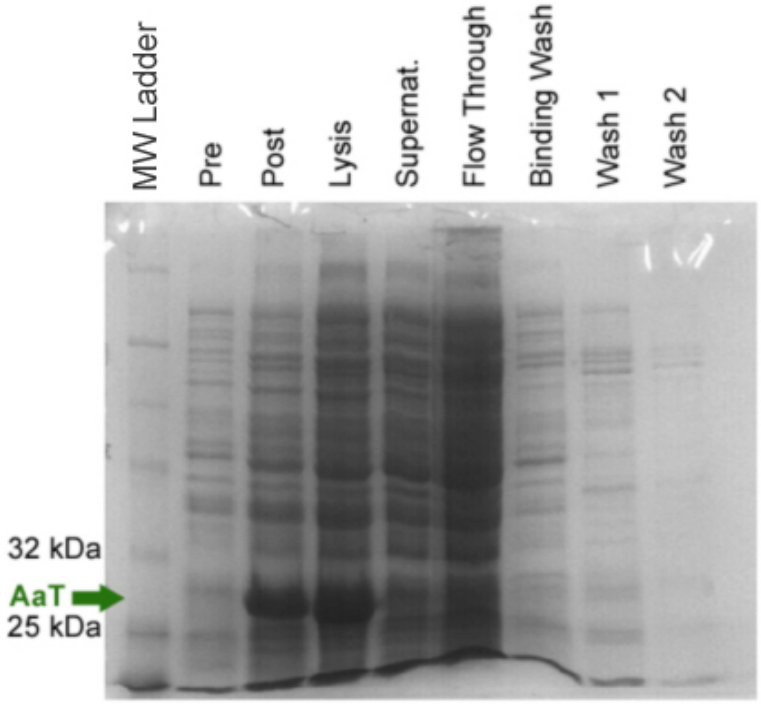


Figure 4-12. SDS PAGE Gel Overexpression Conditions of I_{185}A AaT
Protein was overexpressed and purified from *E. coli* using IPTG overexpression and Nickel Bead purification. Protocol adapted from Graciet *et al*.[50]

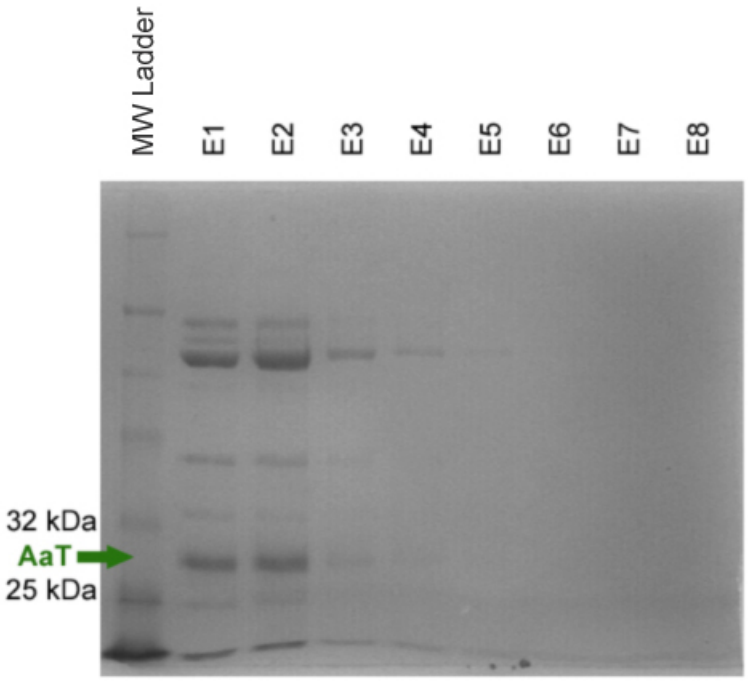


Figure 4-13. SDS PAGE Gel Expression Analysis of Mutant $I_{185}A$ Elution Fractions

Several expression-stable mutants ($M_{144}A$, $L_{170}A$, $I_{185}A$, and $C_{187}A$) were assayed for enzymatic activity with Phe-A and LysAlaAcm to determine a standard of peptide bond-formation activity compared to wild type. For additional comparison, we tested other unnatural aminoacyl adenosine analogs that could be transferred by WT AaT with varying degree: Nap, Azf, and Mef. To complete our assay, we analyzed several of our unnatural amino acids that could not be transferred with WT AaT (Nbd, Mcm, Npt1,8, and Npt2,3) to determine if the mutants could use them as substrates with the "space" made from the amino acid replacements with alanine. Unfortunately, we found that none of the mutants were more active than the WT enzyme toward the substrates that could already be transferred by wild-type, and none could utilize substrates with an R-group larger than napthylalanine (Figure 4-14, Table 4-1).

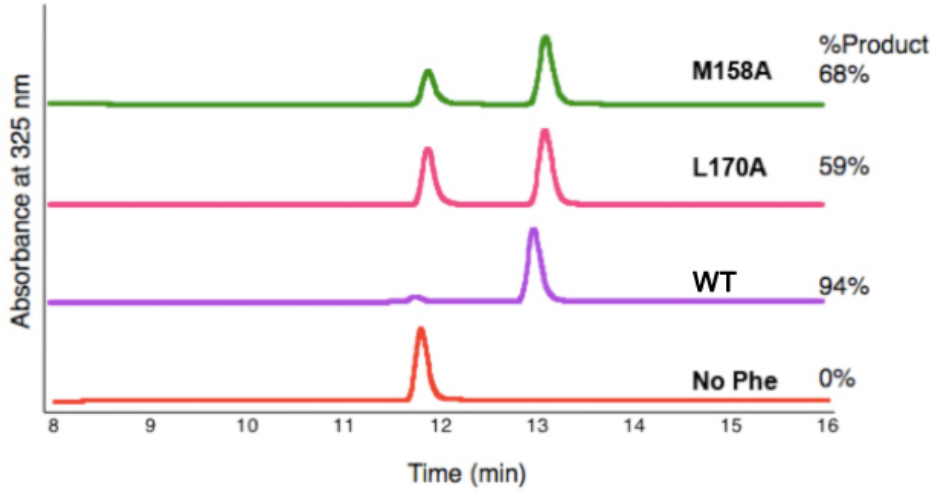


Figure 4-14. Transfer Efficiencies of Phenylalanine by AaT Mutants
Phenylalanine adenosine was ligated onto peptidyl substrate LysAlaAcm *via* AaT and percent ligation was determined by HPLC analysis of peak integration. All reactions were standardized according to the Materials and Methods. The ligation reactions were monitored at 325 nm, the absorption maximum of Acm (aminomethylcoumarin).

Table 4-1. Percent LysAlaAcm Ligation of Analogs Using Mutant AaT

Enzyme	Analog R-Group	% AA-LysAlaAcm[a]
WT AaT	None	0
L170A	Phe	58.6
M158A	Phe	68.0
M144A	Phe	6.2
C187A	Phe	36.4
WT AaT	Phe	98.0
L170A	Nap	74.6
M158A	Nap	32.7
M144A	Nap	27.7
C187A	Nap	3.6
L170A	AzF	41.3
M158A	AzF	35.2
M144A	AzF	18.4
C187A	AzF	4.7
L170A	Nbd	0
M158A	Nbd	0
M144A	Nbd	0
C187A	Nbd	0
WT AaT	Nbd	0
L170A, buffer pH 6	Nbd	0
M158A, buffer pH 6	Nbd	0
WT AaT, buffer pH 6	Nbd	0
L170A	Mef	0
M158A	Mef	0
M144A	Mef	0
C187A	Mef	0
L170A	Mcm	0
M158A	Mcm	0
WT AaT	Mcm	0
L170A	Npt1,8	0[b]
M158A	Npt1,8	0[b]
WT AaT	Npt1,8	0[b]
L170A	Npt2,3	0
M158A	Npt2,3	0
M144A	Npt2,3	10.3[b]
C187A	Npt2,3	0

[a] AA corresponds to the amino acid transferred to the LysAlaAcm peptide N-terminus.
[b] Additional peaks appear in the HPLC spectra but none correspond to product formation.

Additionally, we had some problems in resolving the signal from the fluorescent analogs from the LysAlaAcm product peptide on the HPLC chromatogram (Figure 4-15). Nbd, Npt1,8, Npt2,3, and Mcm all absorb in 325 nm, the wavelength used to monitor Acm. For Nbd, we observed a peak in the HPLC analysis of the LysAlalAcm reaction at 10.9 min, corresponding to the analog. Npt2,3 had the greatest "bleed-through" absorption at 325 nm with a total of three visible peaks. The co-absorptions at 325 nm made for difficult analysis *via* HPLC, but product formation could generally be resolved by MALDI MS analysis of HPLC fractions.

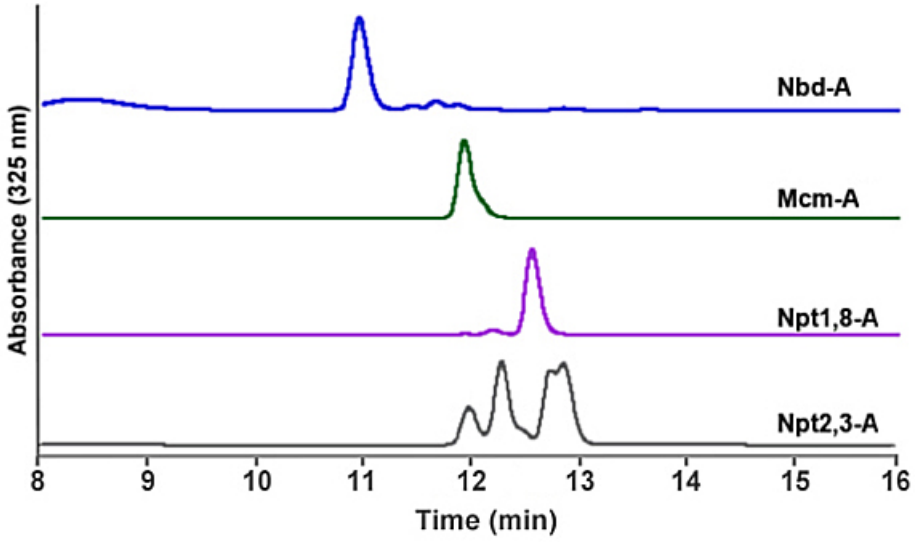


Figure 4-15. Co-absorption of Analogs at 325 nm for LysAlaAcm Ligation HPLCs

We concluded that despite our attempts, the mutants were not as enzymatically efficient or structurally stable as the respective WT AaT. Therefore, we collaborated with the Saven group to design an AaT mutant that, through predictions from computer-aided protein design, would be mutated to become more specific for an unnatural aminoacyl adenosine substrate. Enhanced substrate specificity for a cysteine or

homocysteine disulfide analog would be advantageous because currently our best analog is only at 68 % product formation for *S*-isopropyl cysteine (Csp). Structural changes and amino acid interactions must take place for the accommodation of the disulfide analogs, which have increased polarity and length compared to the natural residue substrates of AaT, phenylalanine and leucine (Figure 4-16). To help understand the unnatural amino acid fit into the AaT active site, the Saven group determined which locations and residue mutations would favorably interact with the respective aminoacyl adenosine analog of interest (Table 4-2). Currently, we have focused our efforts on making a more specific and stable AaT for a variety of substrates, in particular, cysteine disulfides. We have designed primers for the following mutants: $L_{67}M$, $F_{173}L$, $T_{194}A$, and $T_{194}I$. With the help of Haviva Garrett, we have successfully made the following mutants, $L_{67}M$ and $T_{194}A$, and have also overexpressed the $T_{194}I$ mutant. Our efforts are ongoing to make a better mutant AaT that can maintain solubility.

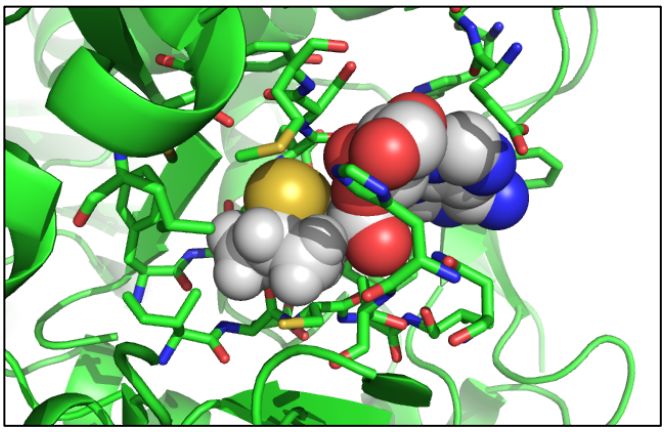


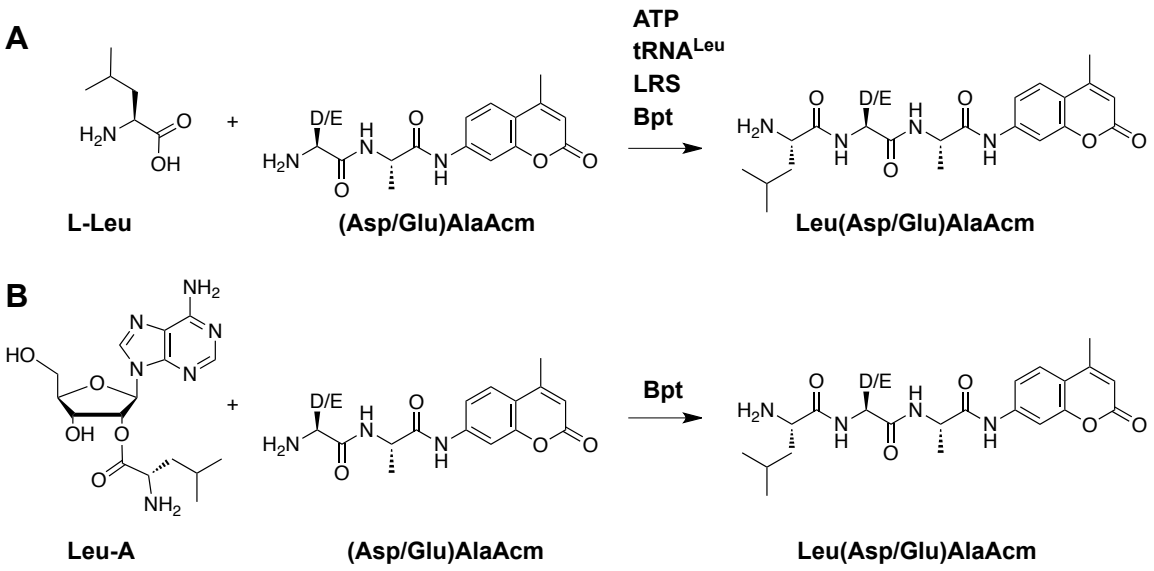
Figure 4-16. Csp-A Docking in WT AaT Active Site
Image was produced using Spartan and the AaT crystal structure (PDB ID: 2Z3K), which had Phe-A docked in the active site.[43]

Table 4-2. Mutagenesis Determined by Computational Modeling

WT *E. coli* AaT*	**L67**	**L72**	**M144**	**M158**	**L170**	**F173**	**I185**	**C187**	**T194**	**L197**
Leu	L67M	L72	M144	M158	L170	F173L	I185	C187	T194A	L197
Met	L67M	L72	M144	M158	L170	F173L	I185	C187	T194I	L197
Csm	L67M	L72	M144	M158	L170	F173L	I185	C187	T194I	L197
Csp	L67M	L72	M144	M158L	L170V	F173M	I185	C187	T194I	L197
Csb	L67M	L72	M144	M158L	L170V	F173M	I185	C187	T194I	L197
Trp	L67M	L72	M144L	M158	L170	F173L	I185M	C187	T194I	L197

- Modeling completed by L. Lu.
- *This sequence is for WT AaT *E. coli* active site amino acids. The following rows are mutations predicted to aid in the specificity of binding to a particular aminoacyl adenosine

In addition to our studies of the AaT active site, we also investigated the possibility of using an alternate transferase, Bpt from *V. vulnificus*. This transferase has a different N-terminal specificity than AaT. Bpt ligates leucine onto proteins with an N-terminal aspartate or glutamate (Scheme 4-1).[50] We found that Bpt is not as efficient an enzyme as AaT, and even with Leu-tRNA and a leucyl tRNA synthetase, the highest yielding product formation of LeuAspAlaAcm was only 44.5 % compared to WT AaT and Phe-tRNA/PheRS going to completion after 4 h (Figure 4-17). We also investigated the use of a Leu-A analog with Bpt and found that its product yield was even worse, 1.6 % LeuAspAlaAcm. Both ligations, enzymatic or chemoenzymatic, had lower percent ligation when used with an N-terminal glutamate peptide, GluAlaAcm. We hypothesized that Bpt might need a molecular chaperone to enhance its enzymatic activity. We have not rigorously tested this hypothesis, but because we noted that AaT and Bpt are expressed in a gene cluster in *V. vulnificus*, we have carried out Bpt reactions in the presence of AaT. No enhancement of Bpt activity was observed.



Scheme 4-1. Peptide Analysis of Bpt Reaction
(A) fully enzymatic and (B) chemoenzymatic with methylcoumarin reporter peptides. The Asp/GluAlaAcm peptides were synthesized by I. Medina.

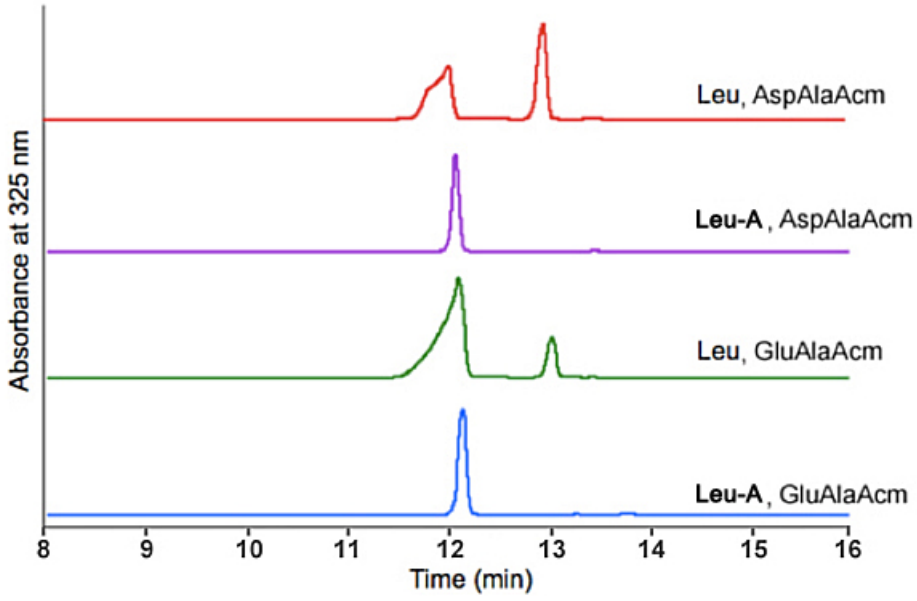


Figure 4-17. HPLC Analysis of Bpt Reaction
Bpt was tested as fully enzymatic (non-analog based) and chemoenzymatic (Leu-A analog used) with two methylcoumarin reporter peptides, AspAlaAcm and GluAlaAcm, monitored at 325 nm.

Through our analysis of AaT, we have concluded that active site mutagenesis is a delicate task. Even one amino acid mutation can make the enzyme insoluble. Therefore, we are hopeful that with the aid of L. Lu of the Saven group that we can rationally-design a more efficient and stable transferase. It is also possible that we will additionally investigate Bpt with *in silico* design to help understand its low degree of enzymatic catalysis *in vitro*. We are continuing our efforts to better understand the mode of enzymatic catalysis and importance of active site residues in AaT.

Section 4.3 – Materials and Methods

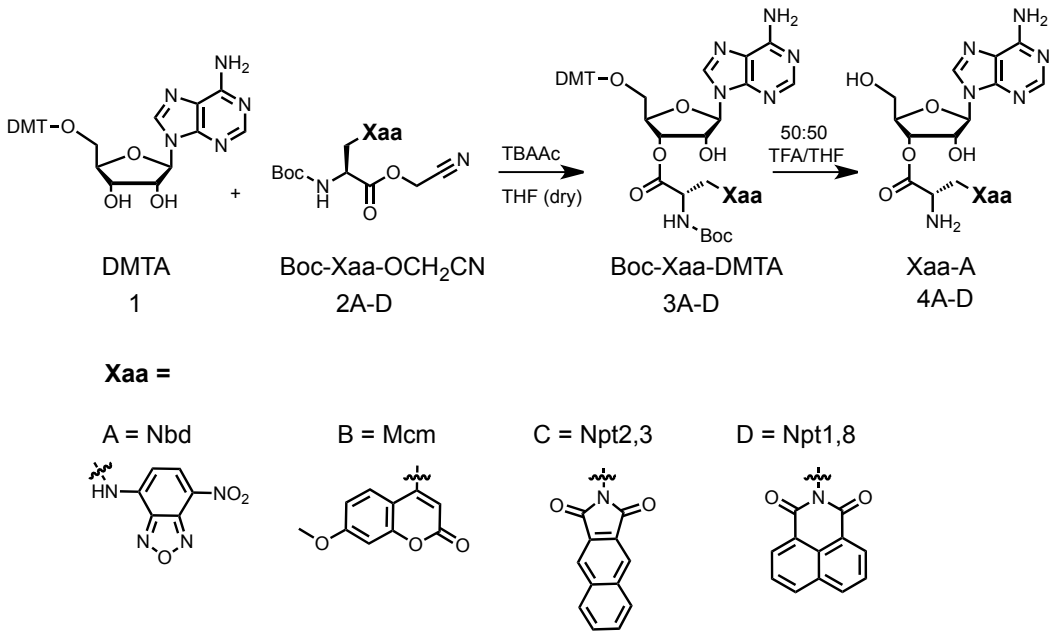
4.3.1 – Materials

General Information. Chloroacetonitrile, *N,N*-diisopropylethylamine (DIPEA), 5′-*O*-(4,4′-Dimethoxytrityl) adenosine ((DMT)-A), 4-Chloro-7-nitrobenzofurazan, tetrabutylammonium acetate (TBAAc), trifluoroacetic acid (TFA), and triisopropyl silane (TIPSH) were purchased from Sigma-Aldrich (St. Louis, MO). (DMT)-A production was discontinued by Sigma-Aldrich, after which it was purchased as a custom order from ChemGenes Corporation (Wilmington, MA) and was additionally synthesized in-house following the protocol outlined by Ogilvie *et al*.[139] Lysylalanyl-aminomethylcoumarin (LysAlaAcm) was purchased from Bachem (Torrance, CA). *N*-Boc-L-phenylalanine (Boc-Phe-OH), 1, 8-naphthalic anhydride and all solvents were purchased from Fisher Scientific (Pittsburgh, PA). 2, 3-naphthalic anhydride was purchased from TCI America (Portland, OR). All deuterated solvents were purchased from Cambridge Isotopes Laboratories, Inc. (Andover, MA). *E. coli* BL21(DE3) cells were purchased from Stratagene (La Jolla, CA). The pEG6 plasmid, containing His_{10}-tagged *E. coli* AaT, was a gift from Alexander Varshavsky (California Institute of Technology). All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). *N*-Boc-L-leucine (Boc-Leu-OH), *N*-Boc-L-*p*-azidophenylalanine (Boc-Azf-OH), *N*-Boc-L-naphthylalanine (Boc-Nap-OH), *N*-Me-L-phenylalanine (Boc-Mef-OH), and *N*-Fmoc-methoxycoumarinylalanine (Fmoc-Mcm-OH) were purchased from Bachem (Torrence, CA). All solvents were purchased from Fisher Scientific (Pittsburgh, PA). QuikChange® site-directed mutagenesis kits were purchased from Stratagene (La Jolla, CA). DNA oligomers were purchased from Integrated DNA Technologies, Inc

(Coralville, IA). Bradford reagent assay kits and protease inhibitor cocktail was purchased from Sigma-Aldrich (St. Louis, MO). All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). Milli-Q filtered (18 MΩ) water was used for all aqueous solutions (Millipore; Billerica, MA). Matrix-assisted laser desorption ionization (MALDI) mass spectra were collected using a Bruker Ultraflex III MALDI-TOF-TOF mass spectrometer (Billerica, MA). UV absorbance spectra were obtained with a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA). Donor molecule purification was conducted on a BioCad Sprint FPLC (GMI Inc.; Ramsey, MN; originally from Perseptive Biosystems) and Varian HPLC with a Waters Sunfire Prep C18-prep OBD column, 5 μm, 17 - 150 mm (Milford, MA). Analytical HPLC assays were performed on an Agilent 1100 HPLC using a Waters Symmetry Shield C18 column. NMR spectra, ^{1}H and ^{13}C, were collected with a Bruker DRX 500 MHz instrument. "Low resolution" electrospray ionization (ESI) mass spectra (LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. DNA sequencing was performed at the University of Pennsylvania DNA sequencing facility.

4.3.2 – Aminoacyl adenosine cyanomethyl ester precursors

The respective aminoacyl adenosine cyanomethyl ester precursors, **2A-D**, were synthesized and characterized by Mark Fegley (Scheme 4-2). Information describing their synthesis and characterization can be found in his Master's degree thesis, M. Fegley, University of Pennsylvania 2007.



Scheme 4-2. Synthesis of Xaa-A Analogs from Cyanomethylester Precursors

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-((7-nitrobenzo[*c*][1,2,5]oxadiazol-4-yl)amino)propanoate (Boc-Nbd-(DMT)-A, 3A).

Tetrahydrofuran (5 mL) was added to Boc-Nbd-OCH_2CN (288.6 mg, 0.710 mmol), (DMT)-A (99.8 mg, 0.175 mmol), and NBu_4Ac (13 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and silica chromatography (100 % ethyl acetate, 2.5 % - 10 % methanol in ethyl acetate), afforded 23.3 mg of an orange solid in 14.5 % yield. R_f 0.5 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR shown below in Figure 4-18; LRMS (ESI) m/z calculated for $C_{45}H_{47}N_{10}O_{12}$ $(M + H)^+$ 919.3, found 919.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-((7-nitrobenzo[*c*][1,2,5]oxadiazol-4-yl)amino)propanoate (Nbd-A, 4A).

Trifluoroacetic acid (1 mL tetrahydrofuran (1 mL) and TIPSH (16.1 mg, 0.021 mL, 0.101 mmol) were added to **3A** (23.3 mg, 0.0254 mmol). The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a final 1 mL water back-extraction against the dichloromethane layer. The water-soluble layer containing **4A** was then HPLC-purified on a C18-prep column using Gradient 1 (Table 4-3). HPLC/MALDI analysis m/z calculated $C_{19}H_{21}N_{10}O_8$ $(M + H)^+$ 517.2; product was found in four separate peaks (17.5 min, 18.3 min, 18.5 min, and 19.3 min), and in all four peaks found 517.1.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(7-methoxy-2-oxo-2*H*-chromen-4-yl)propanoate (Boc-Mcm-(DMT)-A, 3B).

Tetrahydrofuran (5 mL) was added to Boc-Mcm-OCH_2CN (26 mg, 0.0646 mmol), (DMT)-A (35.5 mg, 0.0623 mmol), and NBu_4Ac (4.8 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer chromatography (5 % methanol in ethyl acetate, eluted in 10 % methanol in chloroform), afforded 26.5 mg of a white solid in 44.8 % yield. R_f 0.3 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR shown below in Figure 4-19; LRMS (ESI) m/z calculated for

$C_{49}H_{51}N_6O_{12}$ $(M + H)^+$ 915.4, found 915.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(7-methoxy-2-oxo-2*H*-chromen-4-yl)propanoate (Mcm-A, 4B).

Trifluoroacetic acid (1 mL tetrahydrofuran (1 mL) and TIPSH (18.3 mg, 0.024 mL, 0.116 mmol) were added to **3B** (26.5 mg, 0.0290 mmol). The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a final 1 mL water back-extraction against the dichloromethane layer. The water-soluble layer containing **4B** was then HPLC-purified on a C18-prep column using Gradient 1 (Table 4-3). HPLC/MALDI analysis m/z calculated $C_{23}H_{25}N_6O_8$ $(M + H)^+$ 513.2; retention time 19.9 min, found 513.1; retention time 21 min, found 513.1.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(1,3-dioxo-1*H*-benzo[*f*]isoindol-2(3*H*)-yl)propanoate (Boc-Npt2,3-(DMT)-A, 3C).

Tetrahydrofuran (5 mL) was added to Boc-Npt2,3-OCH_2CN (350.4 mg, 0.828 mmol), (DMT)-A (118.9 mg, 0.207 mmol), and NBu_4Ac (20.9 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and silica chromatography (100 % ethyl acetate, 2.5 % - 10 % methanol in ethyl acetate), afforded 31 mg of an orange solid

in 16 % yield. R_f 0.75 in 5 % methanol in ethyl acetate; ^1H and ^{13}C NMR shown below in Figure 4-20; LRMS (ESI) m/z calculated for $C_{51}H_{50}N_7O_{11}$ $(M + H)^+$ 936.4, found 936.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(1,3-dioxo-1*H*-benzo[*f*]isoindol-2(3*H*)-yl)propanoate (Npt2,3-A, 4C).

Trifluoroacetic acid (1 mL tetrahydrofuran (1 mL) were added to **3C** (31 mg, 0.0331 mmol). The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a final 1 mL water back-extraction against the dichloromethane layer. The water-soluble layer containing **4C** was then HPLC-purified on a C18-prep column using Gradient 1 (Table 4-3). HPLC/MALDI analysis m/z calculated $C_{25}H_{23}N_7O_7$ $(M + H)^+$ 534.2; retention time 23.7 min, found 534.2; retention time 25 min, found 534.3.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl 2-((*tert*-butoxycarbonyl)amino)-3-(1,3-dioxo-1*H*-benzo[*de*]isoquinolin-2(3*H*)-yl)propanoate (Boc-Npt1,8-(DMT)-A, 3D).

Tetrahydrofuran (5 mL) was added to Boc-Npt1,8-OCH_2CN (26.5 mg, 0.0283 mmol), (DMT)-A (15.6 mg, 0.0274 mmol), and NBu_4Ac (2.1 mg,) and stirred for 24 h. The solvent was removed under reduced pressure and preparative thin-layer

chromatography (5 % methanol in ethyl acetate, eluted in 10 % methanol in dichloromethane), afforded 17.2 mg of a white solid in 67.2 % yield. R_f 0.5 in 100 % ethyl acetate; ^1H and ^{13}C NMR shown below in Figure 4-21; LRMS (ESI) m/z calculated for $C_{51}H_{50}N_7O_{11}$ $(M + H)^+$ 936.4, found 936.5.

(*S*)-(2*R*,3*R*,4*R*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-3-yl 2-amino-3-(1,3-dioxo-1*H*-benzo[*de*]isoquinolin-2(3*H*)-yl)propanoate (Npt2,3-A, 4D).

Trifluoroacetic acid (1 mL tetrahydrofuran (1 mL) and TIPSH (11.7 mg, 0.015 mL, 0.074 mmol) were added to **3D** (17.2 mg, 0.0184 mmol). The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a final 1 mL water back-extraction against the dichloromethane layer. The water-soluble layer containing **4D** was then HPLC-purified on a C18-prep column using Gradient 1 (Table 4-3). HPLC/MALDI analysis m/z calculated $C_{25}H_{24}N_7O_7$ $(M + H)^+$ 534.2; retention time 20.5 min, found 534.2; retention time 22 min, found 534.2.

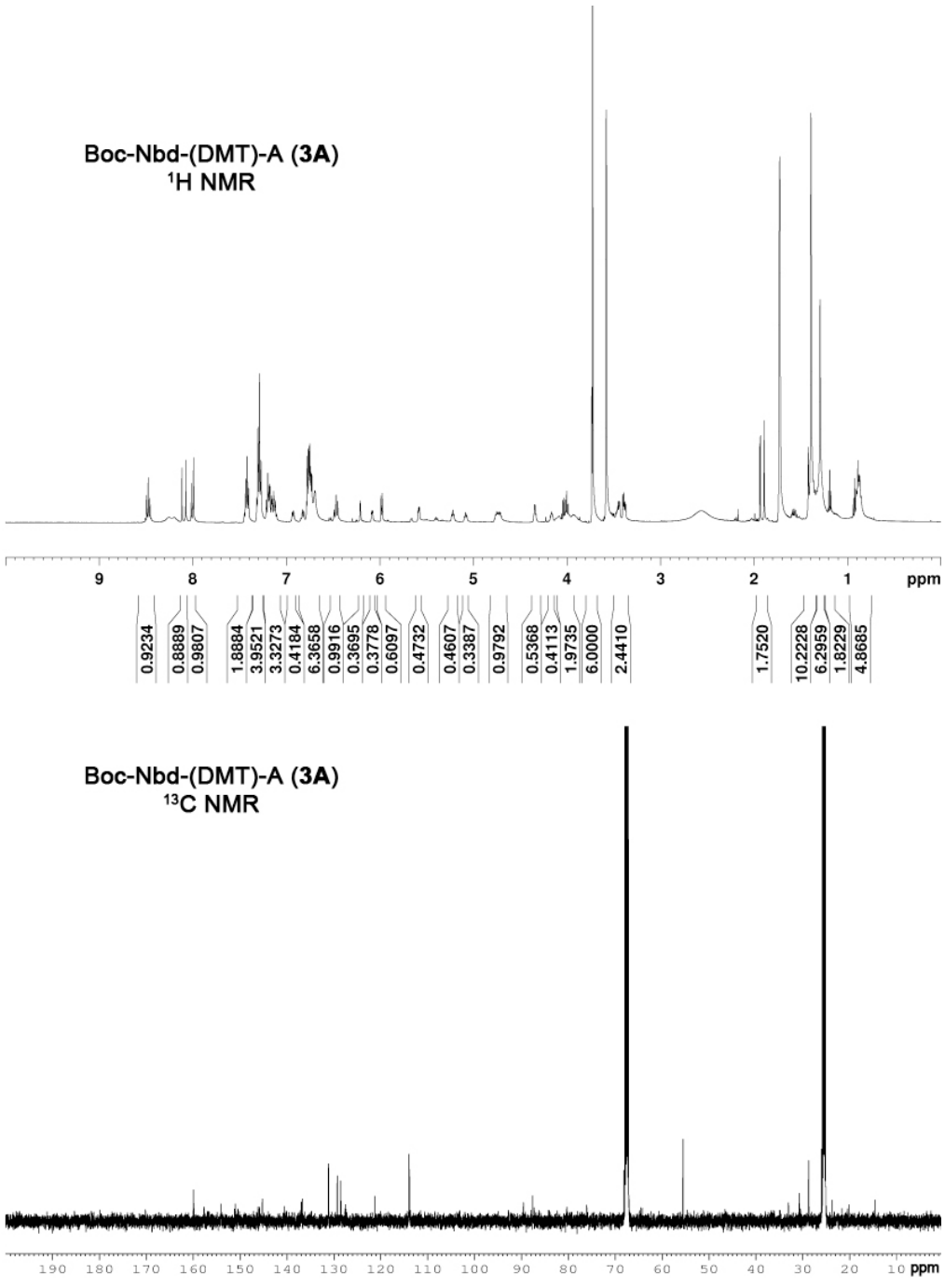


Figure 4-18. ^{1}H and ^{13}C NMR Characterization of Boc-Nbd-(DMT)-A (3A)

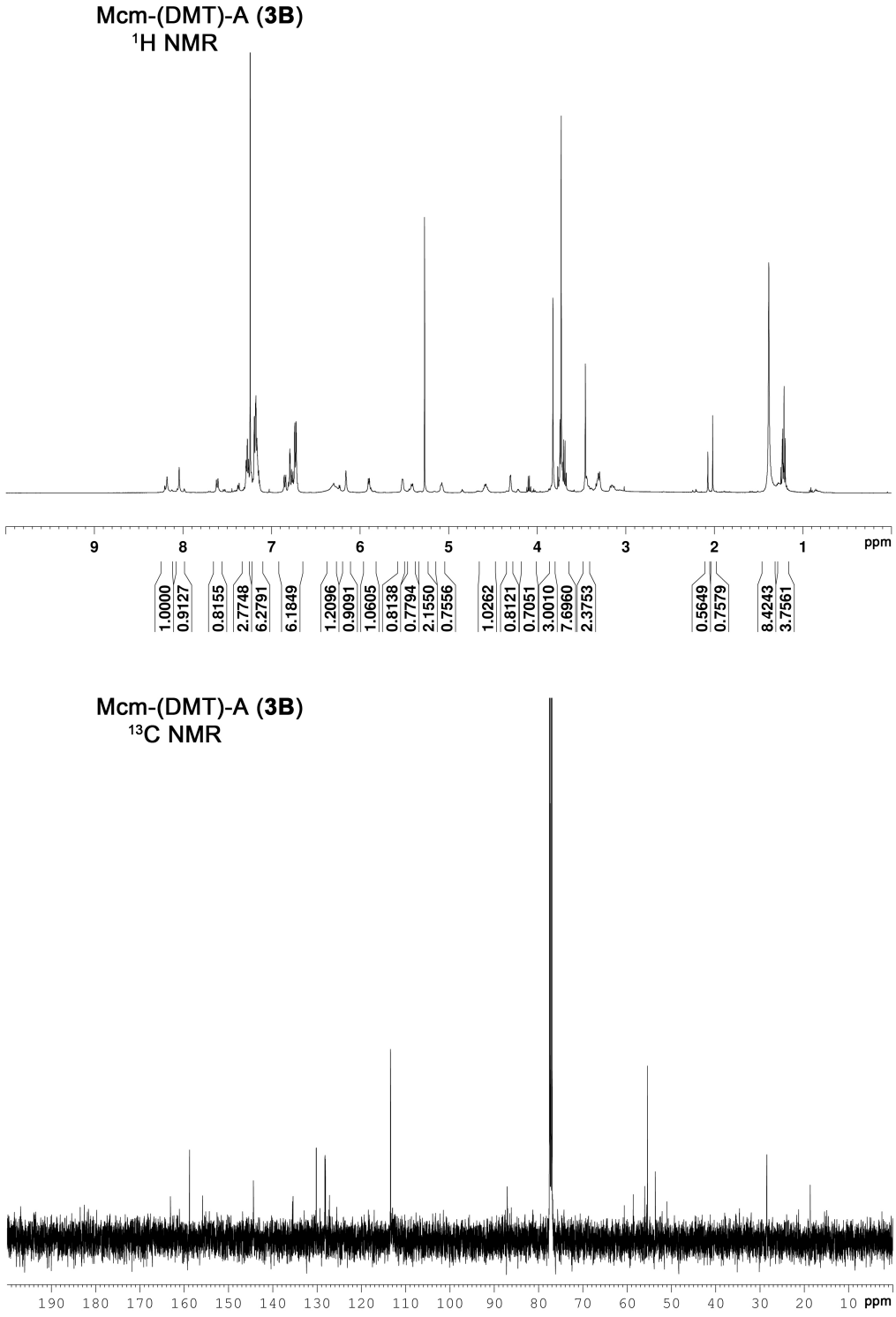


Figure 4-19. ^{1}H and ^{13}C NMR Characterization of Boc-Mcm-(DMT)-A (3B)

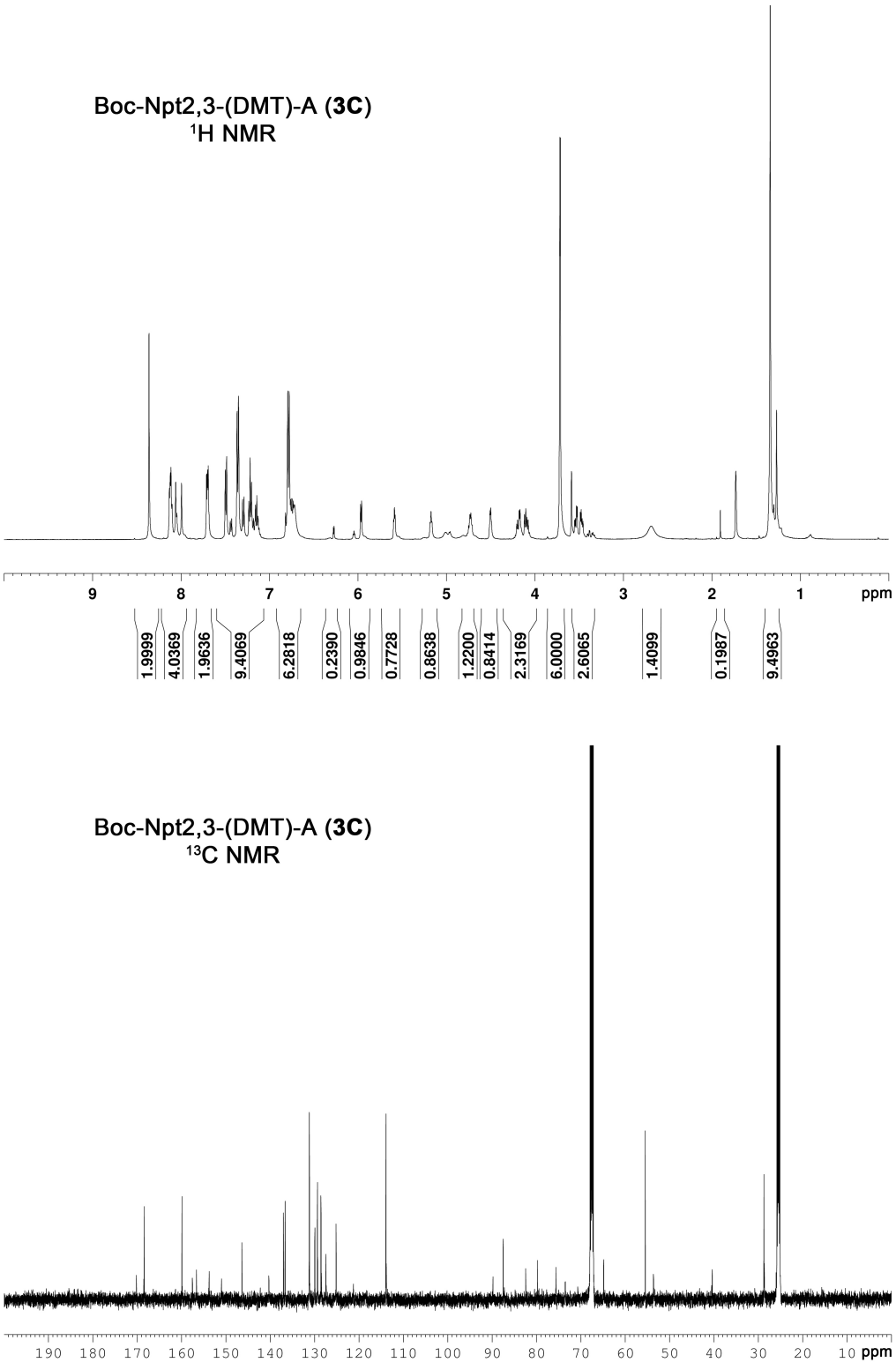


Figure 4-20. ^{1}H and ^{13}C NMR Characterization of Boc-Npt2,3-(DMT)-A (3C)

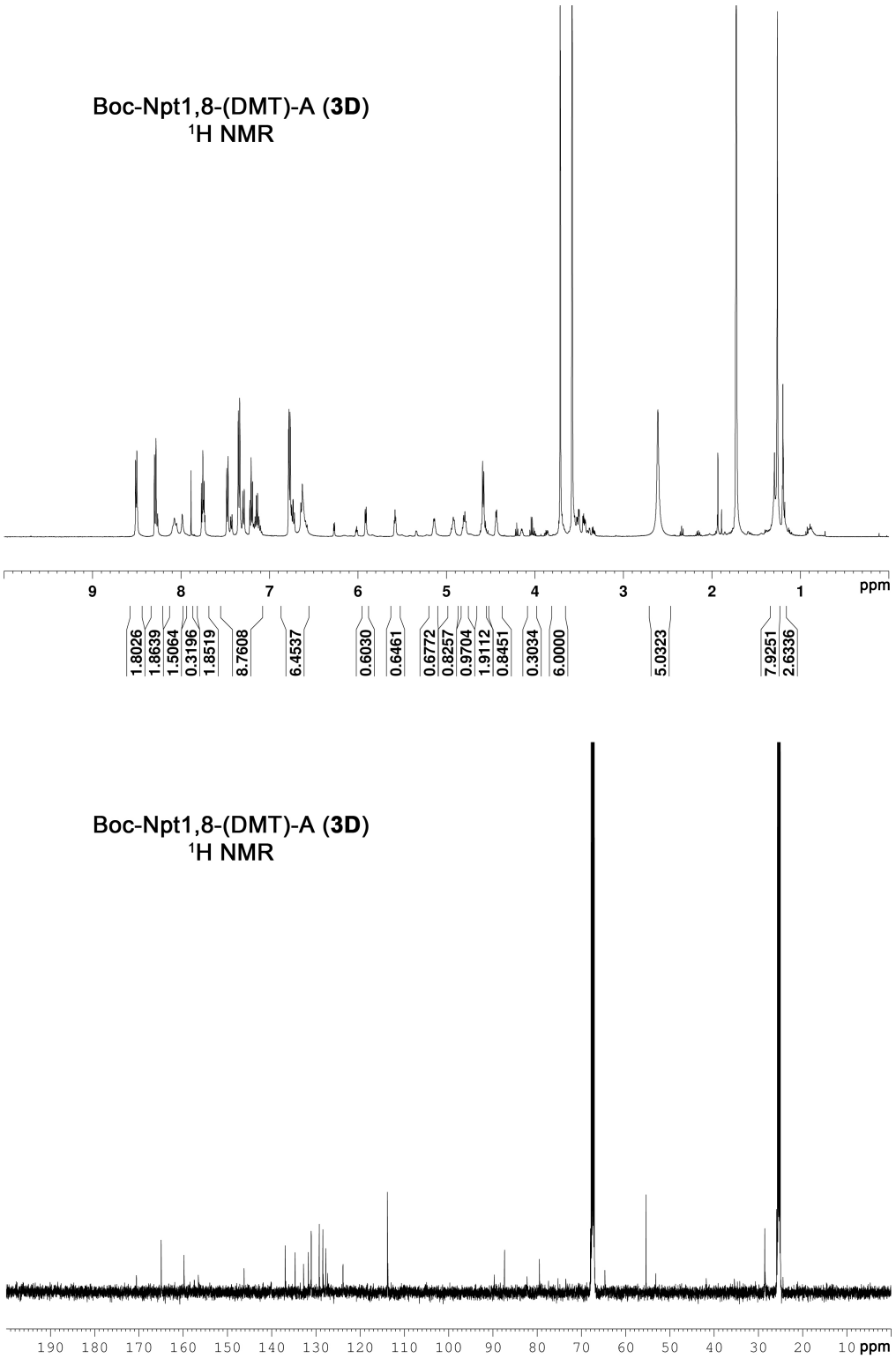


Figure 4-21. ^{1}H and ^{13}C NMR Characterization of Boc-Npt1,8-(DMT)-A (3D)

Table 4-3. Gradients Used for HPLC Purification of Small Molecules and Peptides

Gradient #	Time (min)	Buffer A (%)	Gradient #	Time (min)	Buffer A (%)
1	0:00	99	**2**	0:00	99
15 mL/min	10:00	99	1 mL/min	5:00	99
	40:00	65		10:00	70
	45:00	0		15:00	60
	50:00	0		20:00	0
	55:00	99		25:00	0
	60:00	99		27:00	99
				30:00	99

4.3.3 – Aminoacyl Transferase Expression and Purification

E. coli AaT was expressed from the pEG6 plasmid in *E. coli* BL21-Gold (DE3) cells using a procedure adapted from Graciet *et al*.[50] *E. coli* were grown in a primary culture of 5 mL LB at 37 °C to OD_{600} of 0.5 and then were rediluted into a secondary culture of 500 mL LB and grown to OD_{600} of 0.6. AaT expression was induced using 0.1 mM isopropyl β-D-thiogalactoside and cells were grown at 25 °C for ~16 h. Cells were pelleted at 6, 000 RPM using a GS3 rotor and Sorvall RC-5 centrifuge. Cell pellets were resuspended in the Ni-NTA binding buffer (50 mM Tris, 10 mM imidazole, 300 mM KCl, and 5 mM β-mercaptoethanol, pH 8.0) and included protease inhibitor cocktail, 1 mM PMSF, and 10 units/mL DNAse1–Grade II. Following resuspension, the cells were lysed using sonication. Soluble proteins were collected *via* centrifugation at 13,200 RPM for 15 min. Collected soluble protein was gently shaken for 1 h on ice with Ni-NTA resin. The resin was prepared by rinsing with Ni-NTA binding buffer and then washed with four volumes of Ni-NTA wash buffer (50 mM Tris, 50 mM imidazole, 300 mM KCl, and 5 mM β-mercaptoethanol, pH 8.0). The proteins were eluted with elution buffer (50 mM Tris, 250 mM imidazole, 300 mM KCl, and 5 mM β-mercaptoethanol, pH 8.0). Pure elution fractions of *E. coli* AaT were dialyzed overnight in transferase buffer (50

mM Tris, 30 % glycerol, 120 mM $(NH_4)_2SO_4$, 5 mM β-mercaptoethanol, pH 8.0). The dialyzed enzymes were stored at – 80 °C. Protein concentrations were determined using the Bradford assay and a bovine serum albumin standard curve according to the manufacturer's instructions (Figure 4-22). (Sigma Aldrich; St. Louis, MO)

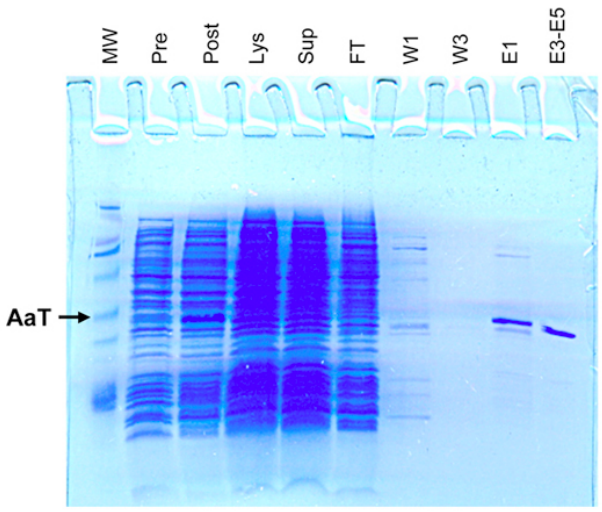


Figure 4-22. SDS PAGE Gel Analysis of AaT Expression and Purification
Lanes (left to right): 1) Molecular weight markers (Masses in kDa: 16, 25, 32, 47, 80, 100, 210); 2) Pre-induction; 3) Post-induction; 4) Crude cell lysate; 5) Supernatant after centrifugation; 6) Ni-NTA column flow-through; 7) Column wash 1; 8) Column wash 3; 9) Elution fraction 1; 10) Combined elution fractions 3-5 after overnight dialysis. Figure reprinted with permission from (Anne M. Wagner, Mark W. Fegley, John B. Warner, Christina L. J. Grindley, Nicholas P. Marotta, and E. James Petersson. *J. Am. Chem. Soc.* **2011**, 133, 15139–15147.). Copyright (2011) American Chemical Society.

4.3.4 – Bpt Transferase Expression and Purification

V. vulnificus Bpt, 29.7 kDa, was expressed from the pEG6 plasmid in *E. coli* BL21-Gold (DE3) cells using the same procedure as discussed above for AaT, which was adapted from Graciet *et al.*[50] The enzyme preparation only differed from AaT in the dialysis buffer. Pure elution fractions of *V. vulnificus* Bpt were dialyzed overnight in 1X Bpt buffer (50 mM Tris, 300 mM potassium chloride, 10 % glycerol, and 15 mM β-mercaptoethanol, pH 8.0). Overexpression of Bpt can be found in Figure 4-23.

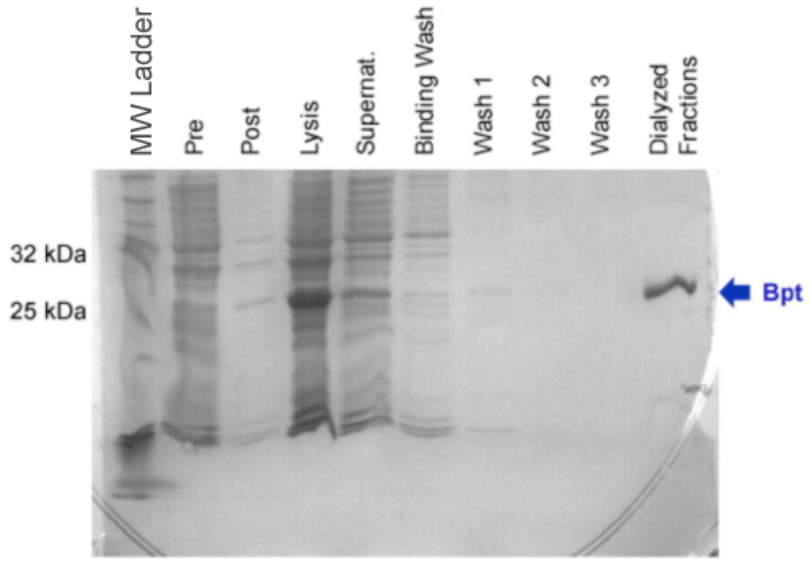


Figure 4-23. SDS PAGE Gel Analysis of Bpt Expression and Purification
Lanes (left to right): 1) Molecular weight markers (Masses in kDa: 16, 25, 32, 47, 80, 100, 210); 2) Pre-induction; 3) Post-induction; 4) Crude cell lysate; 5) Supernatant after centrifugation; 6) Ni-NTA column binding wash flow-through; 7) Column wash 1; 8) Column wash 2; 9) Column wash 3; 10) Combined elution fractions after overnight dialysis.

4.3.5 – Chemoenzymatic LysAlaAcm Ligation Assay

Each ligation reaction, 125 µL total volume, contained the following reagents: aminoacyl adenosine donor (1 mM), recombinant His_{10}-tagged *E. coli* aminoacyl transferase (2.26 µM), and LysAlaAcm (100 µM) in the AaT Ligation Buffer (50 mM HEPES pH 8.0, 150 mM KCl, 10 mM $MgCl_2$). The reaction mixtures were incubated at 37 °C for four hours. At the second, third, and fourth hour of the reaction, an additional 1 mM dose of aminoacyl adenosine donor was added to the reaction mixture. The reaction was quenched with 1 % acetic acid at 4 h. One reaction material modification was made for Hcm analog only, in which we prepared AaT in transferase buffer without the presence of β-mercaptoethanol (50 mM Tris, 30 % glycerol, 120 mM $(NH_4)_2SO_4$). The

proteins were extracted from the reactions *via* acetone precipitation. The reactions were precipitated using 4X reaction volume of acetone and cooled at – 20 °C for 1 h. Next, the reactions were centrifuged at 13,200 rpm at 4 °C for 20 min to separate the reaction from precipitated protein. The supernatant was transferred to fresh 1.5 mL centrifuge tubes and allowed for acetone evaporation overnight at room temperature for all non-disulfide analogs. For all disulfide analogs, the reactions were rotary evaporated immediately to remove acetone in order to minimize disulfide reduction due to air exposure. After acetone evaporation, the supernatant was dried in a Speedvac (Savant, Thermo Scientific, Fisher Inc.) for 30 min to remove residual acetone. The resulting reaction volume was dissolved up to 1.2 mL using MilliQ water and analyzed by HPLC (gradient 1) to determine ligation yield by integration of separated peak intensities monitored at 325 nm (Table 4-4). Collected HPLC fractions were characterized through MALDI MS analysis.

4.3.6 – Fully Enzymatic LysAlaAcm Ligation Assay

Each ligation reaction, 125 µL total volume, contained the following reagents: amino acid (1 mM), ATP (2.5 mM), total *E. coli* tRNA (2 µg/µL), recombinant His-tagged *E. coli* LeuRS, recombinant His_{10}-tagged *E. coli* aminoacyl transferase (2.26 µM), and LysAlaAcm (100 µM) in the AaT Ligation Buffer (50 mM HEPES pH 8.0, 150 mM KCl, 10 mM $MgCl_2$). The reaction mixtures were incubated at 37 °C for 4 h and quenched with 1 % acetic acid. Acetone precipitation and subsequent analysis were performed as described for the Chemoenzymatic LysAlaAcm Ligation Assay.

4.3.7 – AaT Ligation Reaction Analysis by HPLC

All HPLC analyses of LysAlaAcm ligations were monitored on an Agilent HPLC using a Waters C18 column (Milford, MA). The solvents used for peptide purification were the following: 0.1 % trifluoroacetic acid in water (Solvent A) and 0.1 % trifluoroacetic acid in acetonitrile (Solvent B). The HPLC method had the following solvent gradient 1 (Table 4-4). Peptides were monitored at two absorption wavelengths during HPLC analysis, 215 nm for peptide absorption, and 325 nm for Acm absorption. MALDI MS analyses confirmed identity N-terminally modified peptide products (Table 4-5).

Table 4-4. Calculated and Observed masses for LysAlaAcm assay Using Mutant AaT

Product Peptide	Enzyme	$M+H^+$		$M+Na^+$	
		Calc'd	Obs.	Calc'd	Obs.
PheLysAlaAcm	L170A	522.271	522.083	544.254	544.048
PheLysAlaAcm	M158A	522.271	522.098	544.254	544.063
PheLysAlaAcm	M144A	522.271	522.290	544.254	---
PheLysAlaAcm	C187A	522.271	522.328	544.254	544.299
PheLysAlaAcm	WT AaT	522.271	522.096	544.254	544.062
NapLysAlaAcm	L170A	572.287	572.110	594.269	594.065
NapLysAlaAcm	M158A	572.287	572.135	594.269	594.079
NapLysAlaAcm	M144A	572.287	572.302	594.269	594.302
NapLysAlaAcm	C187A	572.287	572.270	594.269	---
AzfLysAlaAcm	L170A	563.272	563.062	585.255	---
AzfLysAlaAcm	M158A	563.272	537.089[a]	585.255	---
AzfLysAlaAcm	M144A	563.272	563.307	585.255	---
AzfLysAlaAcm	C187A	563.272	563.288	585.255	---
NbdLysAlaAcm	L170A	624.252	---	646.235	---
NbdLysAlaAcm	M158A	624.252	---	646.235	---
NbdLysAlaAcm	M144A	624.252	---	646.235	---
NbdLysAlaAcm	C187A	624.252	---	646.235	---
NbdLysAlaAcm	WT AaT	624.252	---	646.235	---
NbdLysAlaAcm	L170A, pH6	624.252	---	646.235	---
NbdLysAlaAcm	M158A, pH 6	624.252	---	646.235	---
NbdLysAlaAcm	WT AaT, pH 6	624.252	---	646.235	---
MefLysAlaAcm	L170A	536.287	---	558.269	---
MefLysAlaAcm	M158A	536.287	---	558.269	---
MefLysAlaAcm	M144A	536.287	---	558.269	---
MefLysAlaAcm	C187A	536.287	---	558.269	---
McmLysAlaAcm	L170A	620.271	---	642.254	---
McmLysAlaAcm	M158A	620.271	---	642.254	---
McmLysAlaAcm	WT AaT	620.271	---	642.254	---
Npt1,8LysAlaAcm	L170A	641.272	---	663.254	---
Npt1,8LysAlaAcm	M158A	641.272	---	663.254	---
Npt1,8LysAlaAcm	WT AaT	641.272	---	663.254	---
Npt2,3LysAlaAcm	L170A	641.272	---	663.254	---
Npt2,3LysAlaAcm	M158A	641.272	---	663.254	---
Npt2,3LysAlaAcm	M144A	641.272	641.283	663.254	---
Npt2,3LysAlaAcm	C187A	641.272	---	663.254	---

[a] Loss of N_2 in MALDI analysis. This has been seen in many Azf-peptide samples.

Table 4-5. Elution Times for LysAlaAcm Ligation Reactions

Enzyme	Analog R-Group	KACm (min)	Product (min)
WT AaT	None	11.03	---
L170A	Phe	11.92	13.13
M158A	Phe	11.92	13.13
M144A	Phe	12.06	13.28
C187A	Phe	12.06	13.28
WT AaT	Phe	11.93	13.14
L170A	Nap	11.93	14.22
M158A	Nap	11.91	14.20
M144A	Nap	11.90	14.19
C187A	Nap	11.90	14.20
L170A	AzF	11.91	13.76
M158A	AzF	11.90	13.75
M144A	AzF	11.89	13.76
C187A	AzF	11.88	13.76
L170A	Nbd	11.91	---
M158A	Nbd	11.92	---
M144A	Nbd	12.06	---
C187A	Nbd	12.06	---
WT AaT	Nbd	11.92	---
L170A, buffer pH 6	Nbd	11.91	---
M158A, buffer pH 6	Nbd	11.90	---
WT AaT, buffer pH 6	Nbd	11.91	---
L170A	Mef	11.90	---
M158A	Mef	11.90	---
M144A	Mef	11.91	---
C187A	Mef	11.91	---
L170A	Mcm	11.97	---
M158A	Mcm	11.96	---
WT AaT	Mcm	11.95	---
L170A	Npt1,8	11.90	---
M158A	Npt1,8	11.90	---
WT AaT	Npt1,8	11.91	---
L170A	Npt2,3	11.88	---
M158A	Npt2,3	11.88	---
M144A	Npt2,3	11.90	12.95*
C187A	Npt2,3	11.89	---

* See co-absorption of analog at this time point, but there is also a small MALDI hit for product.

4.3.8 – Design of an AaT Mutant Library

Quikchange ® mutagenesis was used to generate the following alanine mutations. *E. coli* AaT_{H10}, from the pEG6 plasmid, was mutated at specific residues listed in Tables 4-6 and 4-7. For the double alanine mutants, if the residues were close, we opted to make a double mutant primer, Table 4-7. For all other double mutant combinations, we performed two Quikchange ® mutagenesis steps. The primers used for the collaboration with the Saven group can be found in Table 4-8. All mutagenesis steps were confirmed by DNA sequencing using the T7 promoter primer.

Table 4-6. Single Codon Reassignment Primers for AaT Alanine Mutations

	M144A
(A) M144A	Forward 5' – GCTTGTCGGCGGT**GCG**TACGGCGTGGCC – 3'
	Reverse 5' – GGCCACGCCGTACGCACCGCCGACAAGC – 3'
	M158A
(B) M158A	Forward 5' – GTGGCGAGTCC**GCG**TTCAGCCGGATGGAAAATG – 3'
	Reverse 5' – CATTTTCCATCCGGCTGAACGCGGACTCGCCAC – 3'
	L170A
(C) L170A	Forward 5' – GAAAATGCGTCTAAAACGGCG**GCG**CTGGTATTCTGTGAGG – 3'
	Reverse 5' – CCTCACAGAATACCAGCGCCGCCGTTTTAGACGCATTTTC – 3'
	F173A
(D) L173A	Forward 5' – GCTTCTGGTA**GCG**TGTGAGGAATTTATCGGTCATGG – 3'
	Reverse 5' – CCATGACCGATAAATTCCTCACACGCTACCAGAAGC – 3'
	F177A
(E) L177A	Forward 5' – CTGGTATTCTGTGAGGAA**GCG**ATCGGTCATGGCGGTAAG – 3'
	Reverse 5' – CTTACCGCCATGACCGATCGCTTCCTCACAGAATACCAG – 3'
	I185A
(F) I185A	Forward 5' – CATGGCGGTAAGCTT**GCG**GACTGCCAGGTCCTTAAC – 3'
	Reverse 5' – GTTAAGGACCTGGCAGTCCGCAAGCTTACCGCCATG – 3'
	C187A
(G) C187A	Forward 5' – GCGGTAAGCTTATCGAC**GCG**CAGGTCCTTAACGATCACAC – 3'
	Reverse 5' – GTGTGATCGTTAAGGACCTGCGCGTCGATAAGCTTACCGC – 3'

Table 4-7. Double Codon Reassignment Primers for AaT Alanine Mutations

(A) L170A F173A	Forward 5' – GCGTCTAAAACGGCG**GCG**CTGGTA**GCG**TGTGAGGAATTTATCGG – 3' (L170A, F173A)
	Reverse 5' – CCGATAAATTCCTCACACGCTACCAGCGCCGCCGTTTTAGACGC – 3'
(B) L170A F177A	Forward 5' – CGTCTAAAACGGCG**GCG**CTGGTATTCTGTGAGGAA**GCG**ATCGGTCATGGC – 3' (L170A, F177A)
	Reverse 5' – GCCATGACCGATCGCTTCCTCACAGAATACCAGCGCCGCCGTTTTAGACG – 3'
(C) F173A F177A	Forward 5' – GCGCTTCTGGTA**GCG**TGTGAGGAA**GCG**ATCGGTCATGGCGGTAAGC – 3' (F173A, F177A)
	Reverse 5' – GCTTACCGCCATGACCGATCGCTTCCTCACACGCTACCAGAAGCGC – 3'
(D) I185A C187A	Forward 5' – CGGTCATGGCGGTAAGCTT**GCG**GAC**GCG**CAGGTCCTTAACGATCAC – 3' (I185A, C187A)
	Reverse 5' – GTGATCGTTAAGGACCTGCGCGTCCGCAAGCTTACCGCCATGACCG – 3'

Table 4-8. Codon Reassignment Primers for AaT Mutations for Saven Collaboration

(A) L67M	Forward 5' – GATCCCCGCGCGGTG**ATG**TGGCCAGAATCACTG – 3' (L67M)
	Reverse 5' – CTAGGGGCGCGCCACTACACCGGTCTTAGTGAC – 3'
(B) F173L	Forward 5' – CTAAAACGGCGCTTCTGGTA**CTG**TGTGAGGAATTTATCGGT – 3' (F173L)
	Reverse 5' – GATTTTGCCGCGAAGACCATGACACACTCCTTAAATAGCCA – 3'
(C) T194A	Forward 5' – GGTCCTTAACGATCAC**GCA**GCATCGCTTGGTGC – 3' (T194A)
	Reverse 5' – CCAGGAATTGCTAGTGCGTCGTAGCGAACCACG – 3'
(D) T194I	Forward 5' – CTGCCAGGTCCTTAACGATCAC**ATT**GCATCGCTTGGT – 3' (T194I)
	Reverse 5' – CCATGACCGATAAATTCCTCACACGCTACCAGAAGC – 3'

CHAPTER 5

ATTEMPTED APPLICATIONS OF LEUCYL/PHENYLALANYL AMINOACYL TRANSFERASE

Section 5.1 – Introduction

Alternate routes for the utilization of AaT were attempted over the course of our research in the Petersson Laboratory. We have developed two collaborations with the Bode Laboratory in the Laboratorium für Organische Chemie at the Eidgenössische Technische Hochschule (ETH) in Zürich, Switzerland and the Kashina Laboratory in the University of Pennsylvania Veterinary Medicine School. In both areas, we made progress in the usage at AaT in different ligation reactions and peptide substrates.

With the Bode group, we investigated the possibility of AaT utilization of *N*-hydroxylamine-phenylalanyl adenosine (HONH-Phe-A). Bode *et al.* determined that an α-hydroxylamine can act as a reactive moiety in the amide bond-forming ligation with an α-ketoacid, called the KAHA ligation.[30] If AaT could transfer an amino acid bearing an α-hydroxylamine, it could be possible to bring this ligation from peptides to the protein domain ligation scale. We envisioned using the hydroxylamine as an orthogonal reactive handle in a three-piece protein ligation strategy that could be prepared in a one-pot synthesis (Figure 5-1). Generation of semi-synthetic proteins would take place in a one-pot, chemically-specific dual ligation reaction allowing insertion of synthetic fragments into proteins while taking maximal advantage of cellular biosynthesis.

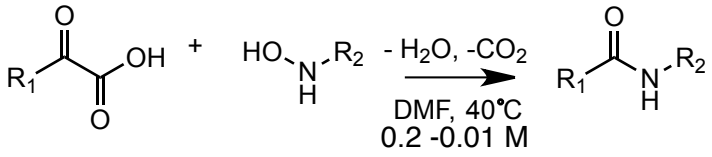


Figure 5-1. KAHA Ligation of a Hydroxylamine and keto-Acid
This reaction, developed by Bode *et al.*, can be done in aqueous conditions with and without substitution at the hydroxylamine oxygen.[30]

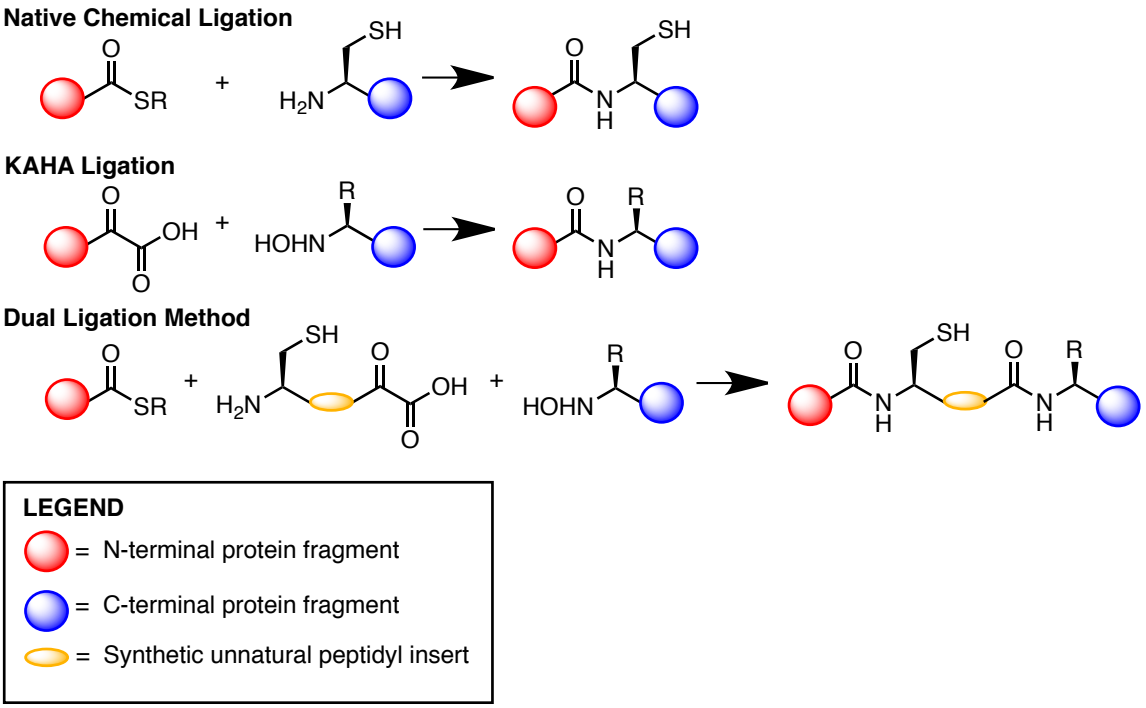


Figure 5-2. KAHA Ligation and NCL Ligation in a One-Pot Protein Dual Ligation Strategy
The reactive handles of each ligation, NCL and KAHA, will only react with their respective partner. The orthogonality and high specificity of these ligations will allow for the reactions to take place in one pot without cross-reactivity between handles.

The Kashina Laboratory is interested in Ate1, a mammalian arginyl transferase. Ate1 is an enzyme that arginylates proteins and is essential for cardiovascular development, actin development and the mammalian N-end protein degradation pathway.[176] The destabilizing residue signals are different for the mammalian pathway, but Ate1 essentially serves the same function as AaT, it ligates a primary destabilizing residue (Arg) onto a secondary destabilizing residue (Asp/Glu) (Figures 5-3, 5-4). One difference from AaT, according to Wang *et al*, is MALDI evidence that Ate1 can also ligate an arginine residue onto aspartates and glutamates that are located internally in addition to the N-terminus.[177] The N-end pathway is only a portion of Ate1 function and the biological function of arginylation is still being elucidated. One way in which the

Kashina lab has learned more about Ate1 arginylation is through a global analysis of mammalian protein substrates that become arginylated using 2D gels and MALDI analysis of *Ate1-/-* and *Ate1+/+* knockout and wild type mouse strains.[178] In conjunction with our research in the Petersson laboratory, we aimed to explore the possibility of using AaT to tag Ate1-arginylated proteins. The goal was to tag N-terminally arginylated proteins with *p*-azidophenylalanine *via* AaT, and then use a click reaction-mediated pull down assay to determine the identity of the arginylated proteins. AaT could possibly identify arginylated proteins that were not found through previous methods.

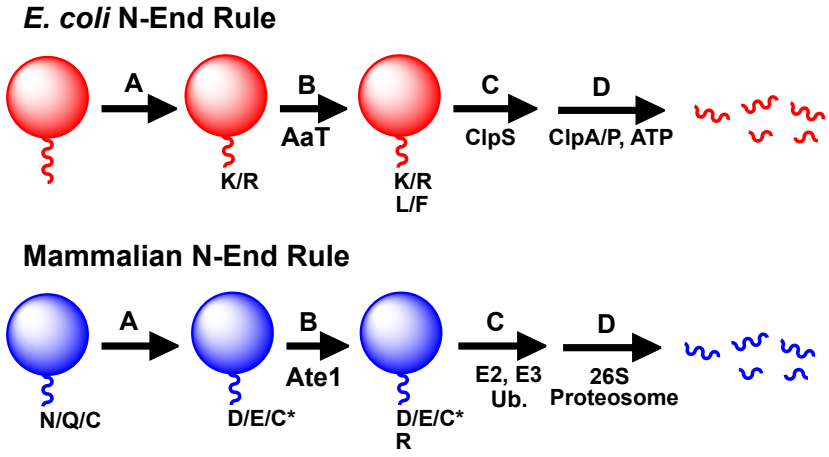


Figure 5-3. Ate1 and AaT Involvement in N-Degradation of Proteins
In the *E. coli* N-End rule a secondary residue, Lys/Arg, is exposed *via* peptidase or exogenous activities (**A**). In **B**, AaT transfers a primary residue, Leu/Phe, to the protein N-terminus. ClpS is a molecular chaperone that recognizes the destabilizing residues (**C**) and brings the protein to the ATP-mediated protein degradation machinery, ClpA/P (**D**). In the mammalian N-End rule, a tertiary destabilizing residues is converted to a secondary destabilizing residue, Asp/Glu/Oxidized Cys, is exposed (**A**). In **B**, Ate1 transfers a primary destabilizing residue, Arg, to the protein N-terminus. E2 and E3 recognize the N-terminal sequence and poly-ubiquitin is transferred to the protein N-terminus (**C**), which signals the protein to be degraded by the 26S proteosome (**D**).[39]

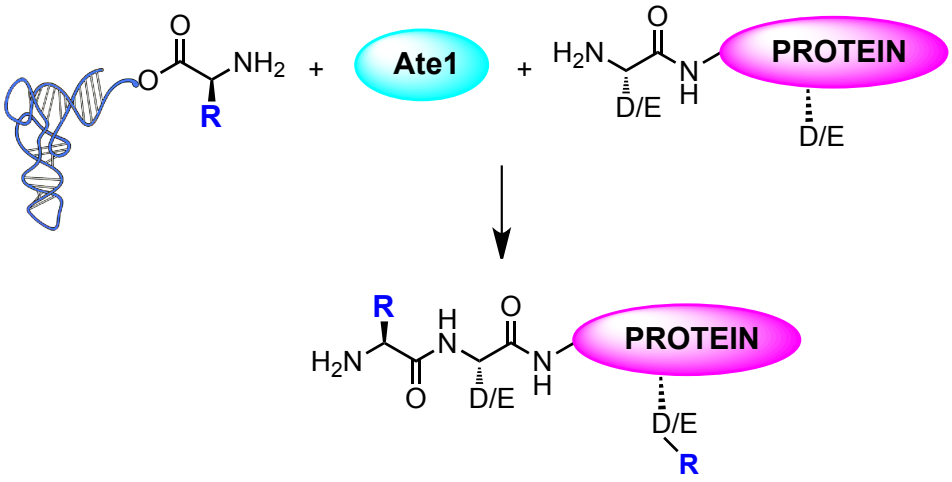


Figure 5-4. Ate1 Reaction Specificity
Ate1 uses arginyl-tRNA as a substrate and transfers arginine onto Asp/Glu at the N-terminus and has been implicated as possibly arginylating at internal Asp/Glu residues.[177]

In this chapter, I will discuss the progress we have made with the Bode and Kashina laboratory collaborations. With the Bode group, our greatest difficulty was the instability of the hydroxylamine, even in its protected form as the phenylnitrone. We thought that we observed transfer of an α-hydroxylamine to the N-terminus of a model peptide, LysAlaAcm, using the analog, *N*-hydroxylamine-phenylalanyl adenosine (HONH-Phe-A); however, we found that we had substantial loss of the hydroxylamine on the product peptide to the α-amine. The instability of the hydroxylamine and the concurrent inability to demonstrate the KAHA ligation on the protein scale (data not shown, determined by I. Medina of the Bode Lab) caused us to abandon this project. For the Kashina laboratory collaboration, the efforts are ongoing, but we have determined that AaT can utilize Ate1-arginylated peptides as substrates. We believe that it will be feasible to bring this multi-enzyme reactivity into cell lysates and possibly to fixed cells.

Section 5.2 – Results and Discussion

For the KAHA ligation, we synthesized two possible substrates for AaT ligation of a hydroxylamine onto the N-terminus of peptides, *N*-hydroxylamine-leucine (HONH-Leu) and HONH-Phe-A (Figure 5-5). We planned to use the HONH-Leu as a free amino substrate that we could use with total synthetase and AaT in a solely enzymatic reaction. The HONH-Phe-A was going to be used as an adenylate small-molecule substrate to be used with AaT without any synthetases present in the reaction. One of the major obstacles of using the HONH-Leu was its insolubility in water requiring it to be dissolved in dimethylsulfoxide. When using this HONH-Leu in a LysAlaAcm transferase reaction, the best product formation was about 52 %; however, according to MALDI, only leucine was transferred to the peptide N-terminus, not leucyl hydroxylamine. Additionally, we attempted to reduce hydroxylamine instability by lowing the pH of the AaT reaction to pH 6.5 and reducing the β-meracaptoethanol concentration in the AaT storage buffer to 5 mM. However, with both changes, again only leucine was ligated to the peptide. We changed our approach to use an adenosyl analog to transfer the α-amino hydroxylamine to avoid the use of a synthetase that could be preferentially selecting the leucine amino acid degraded from the hydroxylamine substrate. One of the issues that we found with the HONH-Phe-A analog again stemmed from hydroxylamine instability again but in the form of its protected precursor *N*-phenylnitrone-Phe-A (See section 5.3, Scheme 2, **10**). The *N*-phenylnitrone-Phe-A degraded while being purified on the HPLC to the hydroxylamine and oxime forms. This degradation effectively negated the last step of the synthesis, deprotection of the hydroxylamine. We used the HPLC purified

hydroxylamine-Phe-A in AaT reactions. Again, we found the same problem observed with the HONH-Leu substrate. The HONH-Phe that was ligated onto the peptide also contained a substantial portion of PheLysAlaAcm peptide, which could not be used as a substrate in the KAHA ligation (Figure 5-6). The 53.1 % product yield was a combination of PheLysAlaAcm peptide with and without the hydroxyl at the α-amine. At the time, the Bode group was unable to perform KAHA ligations under suitable aqueous conditions for subsequent elongation of the transferred HONH-Phe; therefore we decided to halt our research until a more stable hydroxylamine precursor could be synthesized.

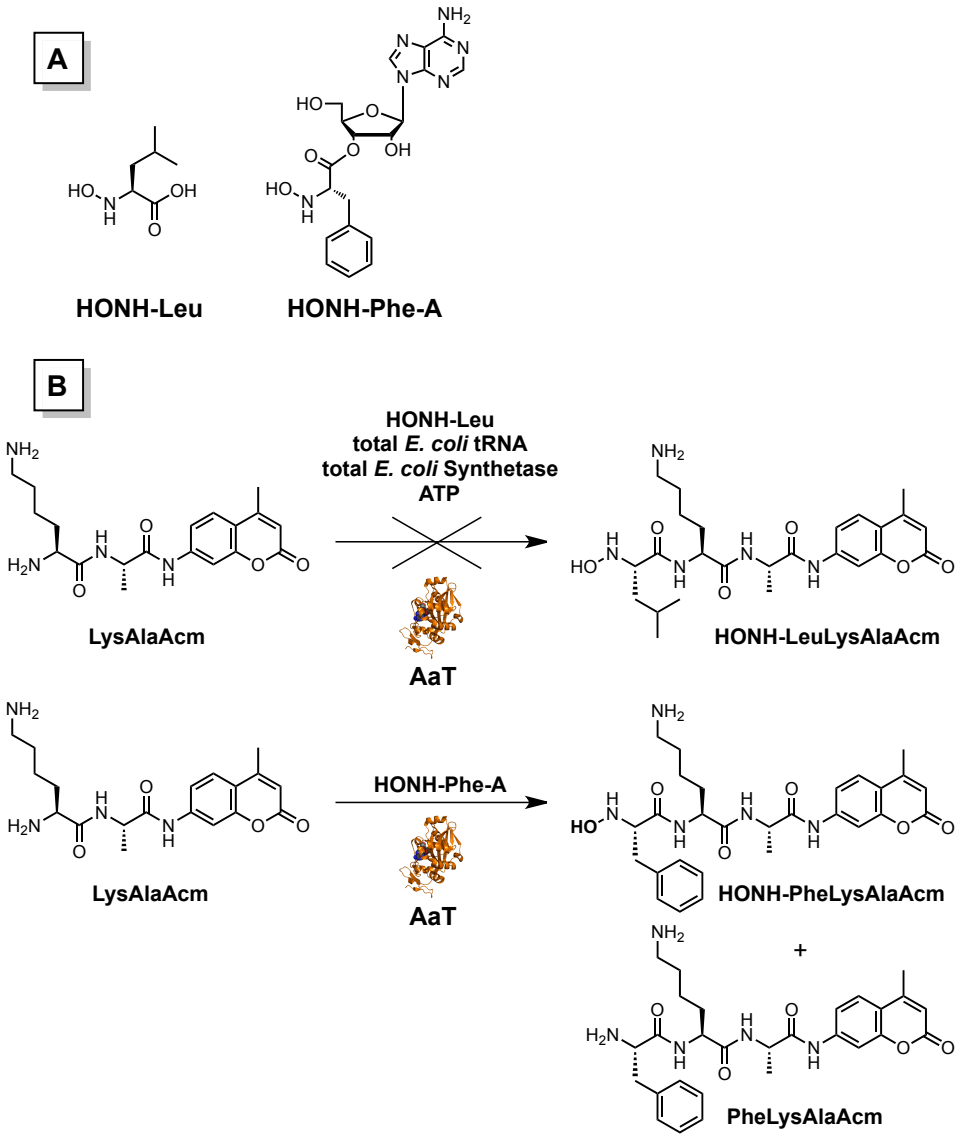


Figure 5-5. AaT Ligation of Hydroxylamine Substrates Onto LysAlaAcm Model Peptide
A. Structures of HONH-Leu and HONH-Phe-A. B. Reactions of hydroxylamine analogs with LysAlaAcm. HONH-Leu could not be used as a substrate for AaT and *E. coli* synthetases. HONH-Phe-A analog could not be transferred without degradation product formation.

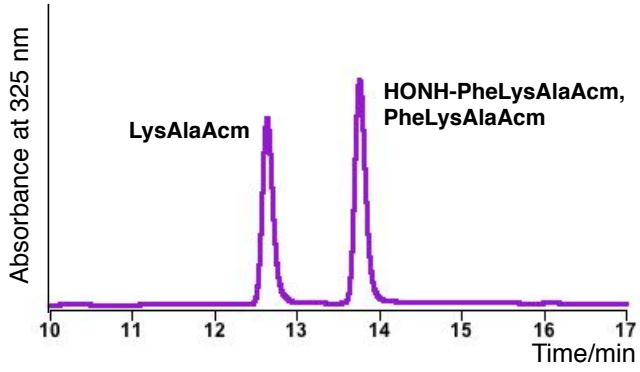


Figure 5-6. HPLC Analysis of AaT ligation of HONH-Phe-A Onto LysAlaAcm Peptide
LysAlaAcm elutes at 12.6 min and HONH-PheLysAlaAcm/PheLysAlaAcm co-elutes at 13.8 min. MALDI confirmation of all HPLC peaks. Co-product percent yield = 53.1 %.

For the Kashina collaboration, we started on the peptide level, and Junling Wang from the Kashina laboratory argninylated two model peptides, AspAlaAcm and GluAlaAcm, synthesized by I. Medina. We found that AaT can use Ate1-arginylated peptides as substrates (Figure 5-7). Using HPLC analysis, we found that the Ate1 reaction produced 63 % ligated product (Figure 5-8). The AaT reaction produced 75 % phenylalanine ligated to the ArgGluAlaAcm peptide. Additionally, we attempted to determine whether Ate1, like AaT, could use minimalist substrates in order to bypass the use of synthetases and arginyl-tRNA substrates. However, Ate1 could not use Arg-A as a substrate. More investigation into minimalist substrates will be required. This research is ongoing and we are investigating the possibility of using this tagging method with *p*-azidophenylalanine in cell lysates.

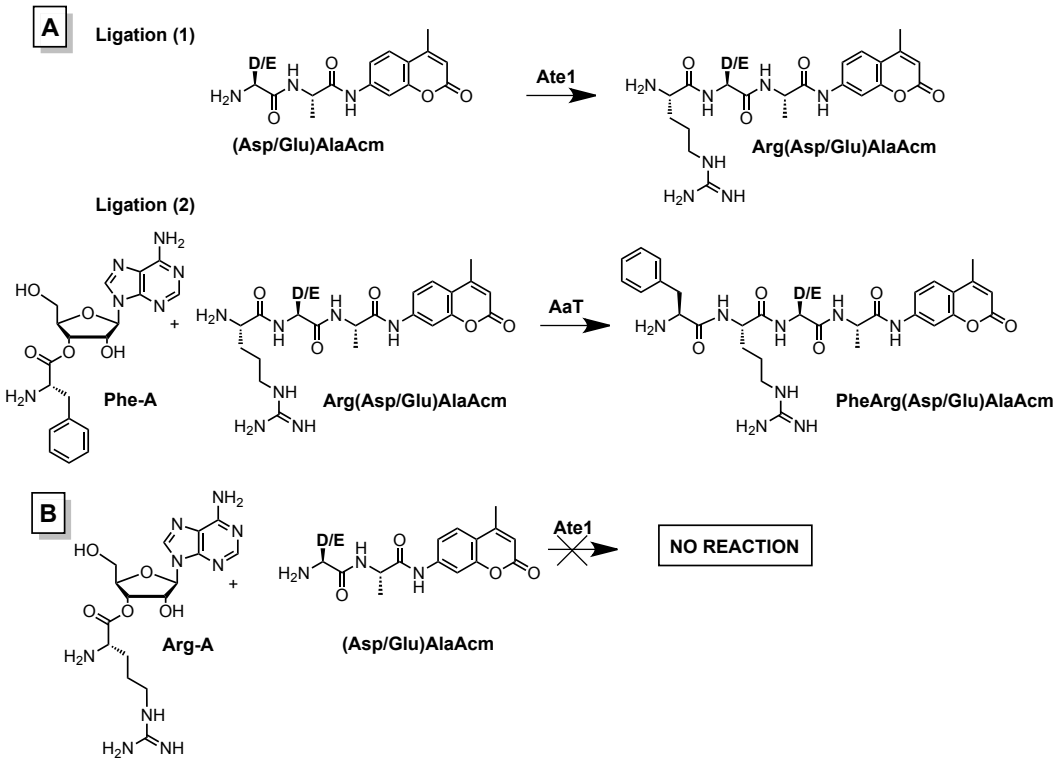


Figure 5-7. Ate1 and AaT Ligation of Peptides

A. Two step N-terminal ligation strategy using Ate1 and AaT. Ligation 1. (Asp/Glu)AlaAcm peptides are ligated via Ate1 (including synthetase, arginyl tRNA, and ATP) with arginine. Ligation 2. Phe-A and arginylated peptides are used as substrates by AaT to form PheArg(Asp/Glu)AlaAcm peptides. B. Arg-A was synthesized but could not be used as a substrate by Ate1.

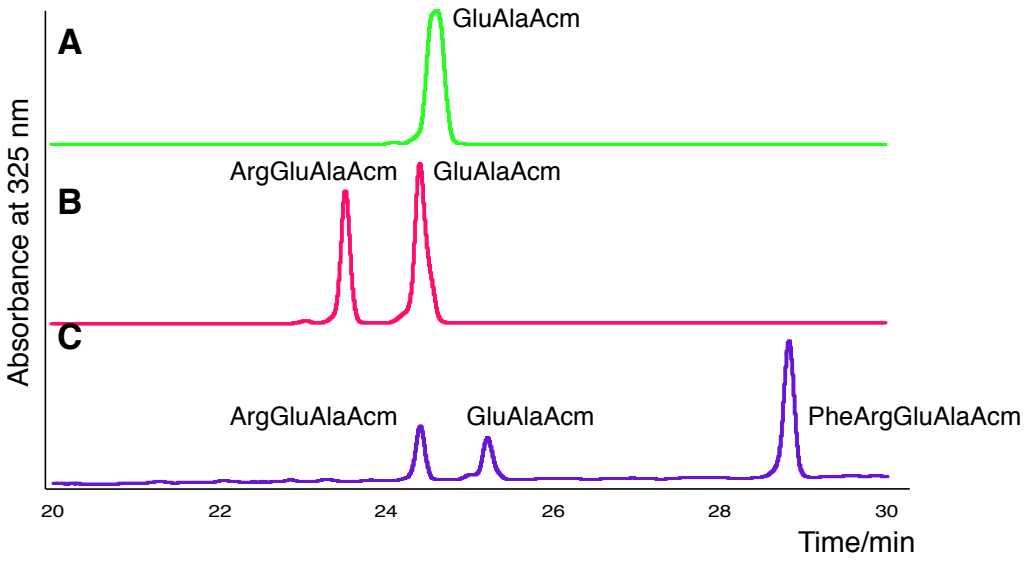


Figure 5-8. HPLC Analysis of Ate1 and AaT Ligation Onto Model Peptides
GluAlaAcm elutes at 24.4 min, ArgGluAlaAcm elutes at 23.5 min, and PheArgGluAlaAcm elutes at 28.8 min. The percent yield of ArgGluAlaAcm was 61 %. The percent yield of PheArgGluAlaAcm was 75 %.

Section 5.3 – Materials and Methods

5.3.1 – Materials

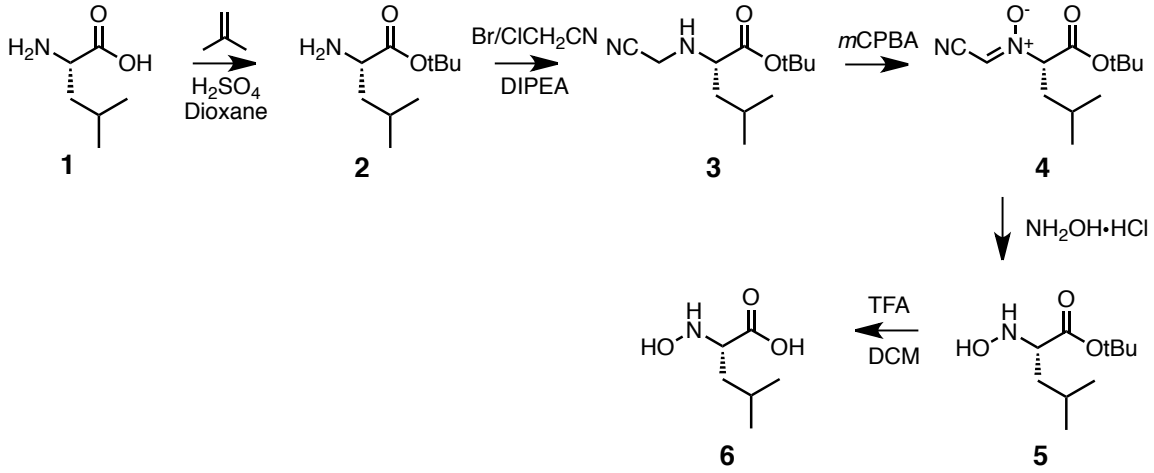
General Information. Chloroacetonitrile, *N,N*-diisopropylethylamine (DIPEA), 5′-*O*-(4,4′-Dimethoxytrityl) adenosine ((DMT)-A), tetrabutylammonium acetate (TBAAc), trifluoroacetic acid (TFA), and triisopropyl silane (TIPSH) were purchased from Sigma-Aldrich (St. Louis, MO). Note: (DMT)-A production was discontinued by Sigma-Aldrich, after which it was purchased as a custom order from ChemGenes Corporation (Wilmington, MA) and was additionally synthesized in-house following the protocol outlined by Ogilvie *et al.*[139] Lysylalanylaminomethylcoumarin (LysAlaAcm) and *N*-Boc-*p*-azidophenylalanine (Boc-Azf-OH) were purchased from Bachem (Torrance, CA). All solvents were purchased from Fisher Scientific (Pittsburgh, PA). All deuterated solvents were purchased from Cambridge Isotopes Laboratories, Inc.

(Andover, MA). *E. coli* BL21(DE3) cells were purchased from Stratagene (La Jolla, CA). The pEG6 plasmid, containing His_{10}-tagged *E. coli* AaT, was a gift from Alexander Varshavsky (California Institute of Technology). All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). Bradford reagent assay kits were purchased from BioRAD (Hercules, CA). Protease inhibitor cocktail was purchased from Sigma-Aldrich (St. Louis, MO). All other reagents were purchased from Fisher Scientific (Pittsburgh, PA). Milli-Q filtered (18 MΩ) water was used for all aqueous solutions (Millipore; Billerica, MA). Matrix-assisted laser desorption/ionization (MALDI) mass spectra were collected with a Bruker Ultraflex III MALDI-TOFTOF mass spectrometer (Billerica, MA). Electrospray ionization (ESI) mass spectra were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA). UV/Vis absorbance spectra were obtained with a Hewlett-Packard 8452A diode array spectrophotometer (Agilent Technologies, Santa Clara, CA).

5.3.2 – Leucyl hydroxylamine (Lha) (6) Synthesis

The synthesis of hydroxylamine donor **6** is given as a general procedure, starting from commercially available Leu-OH (Scheme 1). The carboxy-terminus of the amino acid was first protected with a *t*-butyl group to form **2**. Next, the amino group was activated as a cyanomethyl ester using chloroacetonitrile or bromoacetonitrile to give **3**. Then, the amino group was oxidized using *meta*-Chloroperoxybenzoic acid (*m*CPBA), forming **4**. Then the nitrone hydroxylamine precursor, **4**, was converted to the hydroxylamine, **5**, using hydroxylamine hydrochloride. The carboxyl group was deprotected with trifluoroacetic acid to form the final leucyl hydroxylamine product, **6**.

Scheme 5-1. Hydroxylamine HONH-Leu Synthesis



(*S*)-*tert*-butyl 2-amino-4-methylpentanoate (Leu-O*t*Bu, 2).

Leu-OH (13.12 g, 0.1 mol) was dissolved in 200 mL dioxane in a three-arm round bottom flask. 3.7 equivalents of sulfuric acid (20 mL, 0.37 mol) were added to the reaction dropwise. After dissolution of materials, the round bottom flask was placed in an ice bath with stirring. The reaction was kept under nitrogen and a cold finger and gas condenser were attached to keep the isobutylene cold and liquefied. The gas condenser was filled with 100 mL of isobutylene and released into the round bottom flask; this step was repeated a second time to add a total of 21.4 equivalents (200 mL, 2.14 mols). Reaction was mixed overnight on ice. The reaction was worked up by first neutralizing the acidic conditions with NaOH and then was extracted with ether. The combined organic layers were dried with sodium sulfate (anhydrous). The extract was crystallized as the hydrochloride salt using dry hydrochloric acid (11.6 equiv., 10 mL, 1.6 moles) and afforded 9.96 g, 44.3 %. Product was characterized using NMR: ^{1}H NMR (500 MHz, MeOD) δ 3.91 – 3.88 (m, 1H), 1.82 – 1.76 (m, 2H), 1.68 – 1.62 (m, 1H), 1.54 (s, 9H),

1.04 –1.01 (m, 6H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 167.3, 82.2, 50.3, 38.1, 25.5, 23.0, 19.8. Product does not ionize well on ESI-MS.

(*S*)-*tert*-butyl 2-((cyanomethyl)amino)-4-methylpentanoate (*N*-CH_2CN-Leu-O*t*Bu, 3).

Leu-O*t*Bu, **2**, (1 g, 4.45 mmol) was dissolved in 9 mL of dry acetonitrile in an oven-dried 25 mL round bottom flask. 1.1 equivalents of bromoacetonitrile (0.34 mL, 4.9 mmol) and 2 equivalents of diisopropylethylamine (DIPEA) (1.55 mL, 8.9 mmol) were added to the reaction mixture under nitrogen *via* syringe. The reaction was mixed at room temperature overnight. After evaporation of solvent, the product was purified on silica gel (gradient solvent system: 10 – 20 % ethyl acetate in hexanes to give 0.5 g (49.4 %). R_f 0.5 in 20 % EtOAc in hexanes. ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.72 – 3.58 (m, 2H), 3.28 – 3.23 (m, 1H), 1.80 – 1.75 (m, 2H), 1.50 (s, 9H), 0.94 (t, *J* = 6.25, 6 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 173.9, 117.8, 82.0, 59.6, 42.5, 36.1, 28.2, 24.9, 22.9, 22.2. Product does not ionize well on ESI-MS.

(*S,Z*)-1-(*tert*-butoxy)-*N*-(cyanomethylene)-4-methyl-1-oxopentan-2-amine oxide (*N*-CH_2CN-nitrone-Leu-O*t*Bu, 4).

N-CH_2CN-Leu-O*t*Bu, **3**, (0.5 g, 2.2 mmol) was dissolved in 5.5 mL of dichloromethane under nitrogen in an ice bath. Two equivalents of *m*CPBA (759 mg, 4.4 mmol) were added in small scoops over 30 minutes. After addition of mCPBA, the reaction was mixed for 30 min at room temperature. The reaction was monitored *via* TLC and worked up upon reaction completion, R_f product = 0.5 in 20 % ethyl acetate in

hexanes. To quench the reaction, 5 mL saturated aqueous sodium thiosulfate and 15 mL saturated sodium bicarbonate were added. Slurry was mixed for 15 minutes until all white solid dissolved. Product was extracted with dichloromethane and dried with anhydrous sodium sulfate. The product was characterized using NMR: ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.81 (s, 1H), 4.61 – 4.58 (m, 1H), 2.11 – 2.05 (m, 1H), 1.85 – 1.79 (m, 2H), 1.48 (s, 18.3H), 0.98 – 0.92 (m, 9H). Product was assumed to go to 100% completion and taken directly to the next synthetic step to form the hydroxylamine.

(*S*)-*tert*-butyl 2-(hydroxyamino)-4-methylpentanoate (Lha-O*t*Bu, 5).

N-CH_2CN-nitrone-Leu-O*t*Bu, **4**, (0.53 g, 2.2 mmol) was dissolved in 5.5 mL of methanol. Five equivalents of hydroxylamine hydrochloride (0.76 g, 11 mmol) were added and the reaction was mixed overnight at 60 ºC. R_f = 0.4 in 20 % ethyl acetate in hexanes. Unreacted hydroxylamine hydrochloride was precipitated from the reaction using excess dichloromethane. Aqueous sodium bicarbonate was added and the product was extracted using dichloromethane. Collected organic layers were dried with anhydrous sodium sulfate. 2 equivalents of oxalic acid (0.4 g, 4.4 mmol) dissolved in 2 mL methanol were added to dried extract product. The product was crashed out as a salt using excess ether. No further purification was needed. ^{1}H NMR (500 MHz, DMSO) δ 3.71 (s, 2 H), 3.49 (s, 1 H), 1.42 (s, 9 H), 0.87 (s, 6 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 162.3, 64.3, 37.9, 28.3, 25.2, 23.1, 22.7. NMR had poor resolution. MALDI MS calcd m/z for $C_{10}H_{22}NO_3$ $(M + H)^+$ 204.16, found 203.8.

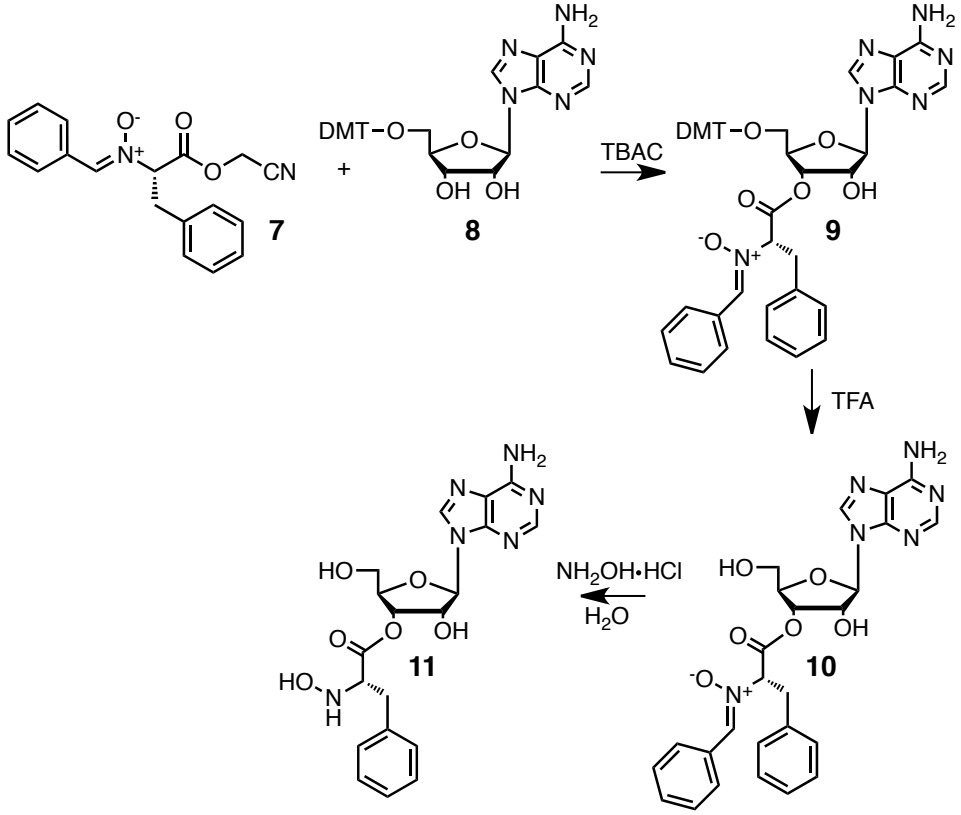
(*S*)-2-(hydroxyamino)-4-methylpentanoic acid (Lha, 6).

Lha-O*t*Bu, **5**, was deprotected with the addition of 1 mL 50 % trifluoroacetic acid and 50 % dichloromethane. 0.5 mL acetonitrile was added to the reaction to help solubilize Lha-O*t*Bu. The reaction was monitored via TLC every 30 min until complete in 1.5 hr. R_f = baseline in 50% ethyl acetate in hexanes. No further purification was needed. ESI MS calcd m/z for $C_6H_{13}NO_3$ $(M + H)^+$ 148.1, found 148.6. NMR had poor resolution and not all peaks were accounted for, but there is a loss of t-butyl peak. ^{1}H NMR (500 MHz, DMSO) δ 2.52 – 2.49 (m, 3), 1.73 – 1.46 (m, 6), 0.92 – 0.87 (m, 6 H).

5.3.3 – Phenylalanyl hydroxylamine adenosine (HONH-Phe-A) (11) Synthesis

The synthesis of Phenylalanyl hydroxylamine adenosine **11** is given as a general procedure, starting from the phenyl nitrone cyanomethyl ester synthesized by I. Medina according to Bode *et al.* (Scheme 2).[30] The protected amino acid was attached to one of the available hydroxyls (2′ OH or 3′ OH) of (DMT)-A, which is commercially available, without preference using a catalytic amount of TBAC. The resulting protected adenylate (**9**) was deprotected using a 50/50 mixture of TFA and THF with TIPSH present as a scavenger, producing **10**. The final product **11** containing the *N*-hydroxylamine, HONH-Phe-A, was formed by addition of hydroxylamine hydrochloride.

Scheme 5-2. Hydroxylamine HONH-Phe-A Analog Synthesis



(*S*,*Z*)-1-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-2-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-3-yl)oxy)-*N*-benzylidene-1-oxo-3-phenylpropan-2-amine oxide (*N*-phenylnitrone-Phe-(DMT)-A, 9).

(DMT)-A (26.2 mg, 0.0460 mmol) was dissolved in 5 mL of tetrahydrofuran (dried with 4 Å molecular sieves) in an oven-dried 4 dram reaction vessel. Twenty equivalents of **7** (267.7 mg, 0.878 mmol) were added to the reaction mixture. TBAAc (27.6 mg, 92 μmol), the catalyst, was added last to the reaction mixture. The reaction was stirred under argon overnight at room temperature overnight. After evaporation of

solvent, the product was purified on silica gel (gradient solvent system: 20 – 0 % petroleum ether in EtOAc, 5 % methanol in EtOAc) to give 24.4 mg (65 %) after evaporation. R_f 0.5 in 5 % methanol in EtOAc. LRMS (ESI) calcd m/z for $C_{47}H_{45}N_6O_8$ $(M + H)^+$ 821.3, $C_{47}H_{44}N_6O_8Na$ $(M + Na)^+$ 843.3, found 821.3, 843.3.

(*S*,*Z*)-1-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-4-hydroxy-2-(hydroxymethyl)tetrahydrofuran-3-yl)oxy)-*N*-benzylidene-1-oxo-3-phenylpropan-2-amine oxide (*N*-phenylnitrone-Phe-A, 10).

9 (12.2 mg) was dissolved in 1 mL of THF and 1 mL trifluoroacetic acid (TFA) and 2.3 mg oxalic acid and reacted for 30 min. The reaction mixture was reduced to dryness by rotary evaporation and extracted using 1 mL dichloromethane and 1 mL water twice with a 1 mL water final back-extraction against the dichloromethane layer. The water-soluble layer containing **10** was then HPLC-purified on a C18-prep column using an increasing acetonitrile gradient (gradient 1 used for Phe-A discussed in Chapter 2 section 2.3). NMR analysis of **10** can be found in Figure 5-9. HPLC/MALDI analysis m/z calcd $C_{26}H_{27}N_6O_6$ $(M + H)^+$ 519.2; retention time 23 - 25 min, found 519.3; retention time 38.1-38.5 min, found 519.3. During the process of HPLC purification, additional side products were found, (1) loss of phenyl nitrone producing hydroxylamine at retention time 21.5 – 25 min ($(M + H)^+$ m/z = 431.2) and (2) formation of oxime at 40.6 min ($(M + H)^+$ m/z = 429.1). Over time in water, product **10** converted into **11**.

(*S*)-(2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-4-hydroxy-2-(hydroxymethyl)tetrahydrofuran-3-yl 2-(hydroxyamino)-3-phenylpropanoate (HONH-Phe-A, 11).

10 (11 mg) was dissolved in to a 0.2 M concentration in methanol (104 μL). 5 equivalents of hydroxylamine hydrochloride was added (7.9 mg, 0.114 mmol) and reacted for 4 h at 60 ºC, monitoring *via* MALDI. At the end of the reaction, mixture was diluted with water and lyophilized to remove solvents. MALDI analysis m/z calcd $C_{19}H_{23}N_6O_6$ $(M + H)^+$ 431.2, found 430.9.

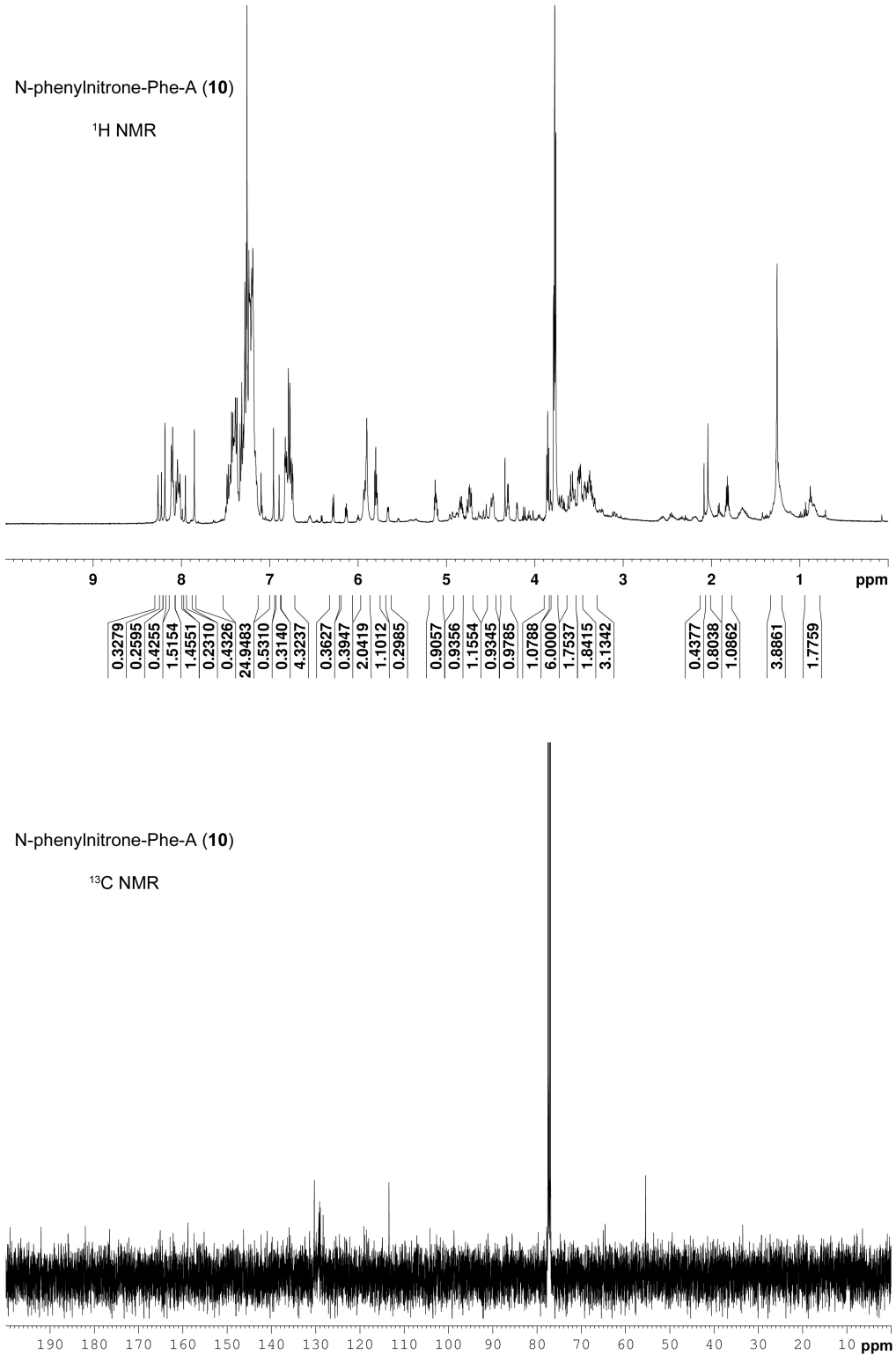


Figure 5-9. ^{1}H and ^{13}C NMR Characterization N-phenylnitrone-Phe-A (10)

5.3.4 – Ate1 and AaT Expression and Purification

E. coli AaT was expressed from the pEG6 plasmid in *E. coli* BL21-Gold (DE3) cells using a procedure adapted from Graciet *et al* and explained in detail in section 2.3.6 of Chapter 2.[50] Ate1 was overexpressed and purified by the Kashina Laboratory.

5.3.5 – Model Peptide Assays for AaT and Ate1

For the AaT peptide ligations, they were performed the same as previously described in Ch. 2 section 2.3.7. All of the AaT ligation reactions using hydroxylamine substrates were purified by HPLC using the same method as described in Chapter 2 section 2.3.10. Ate1 arginylation reactions of peptides Asp/GluAlaAcm were performed by the Kashina laboratory according to standard published procedures.[177] All HPLCs for Asp/GluAlaAcm peptides and ligations were purified using the method outlined in Table 5-1. Collected HPLC fractions were characterized through MALDI MS analysis (Table 5-2).

Table 5-1. Gradient Used for HPLC Purification of Asp/GluAlaAcm Peptides and Ligations

Gradient #	Time (min)	Buffer A (%)
1 mL/min	0:00	99
	5:00	99
	30:00	74
	35:00	0
	40:00	0
	45:00	99

Table 5-2. HPLC Analysis of (Asp/Glu)AlaAcm Transfer Reactions

Product Peptide	Retention Time (min)	$M+H^+$		$M+Na^+$	
		Calc'd	Obs.	Calc'd	Obs.
AspAlaAcm	24.0	362.135	362.395	384.117	---
ArgAspAlaAcm	22.9	518.236	518.462	540.218	540.418
PheArgAspAlaAcm	**	665.305	665.515	687.287	687.495
GluAlaAcm	24.4	376.151	376.429	398.133	398.398
ArgGluAlaAcm	23.5	532.252	532.487	554.234	554.480
PheArgGluAlaAcm	28.8	679.320	679.482	701.302	701.458

**This reaction was not HPLC purified.

CHAPTER 6

CONCLUSIONS

Current research has shown that there are a variety of ways to modify proteins, depending on your reaction restrictions. We have found a new and versatile chemoenzymatic way to modify proteins that fulfills a wide variety of constraints: site-specificity, amide bond formation at the ligation site, amino acid substrate variety for both peptide and small molecule substrates, aqueous conditions, near physiological pH, high salt concentrations, in the presence of denaturants, and with high reproducibility. The AaT amide bond forming reaction is slow, but the native N-terminal specificity outweighs its unfavorable kinetics. We have manipulated a natural enzyme to accept and ligate a variety of small molecule substrates that range from *p*-azido-phenylalanine to protected disulfides without the need for mutagenesis. We hope through the aid of *in silico* studies with the Saven group, we will be able to target an aminoacyl adenosine substrate and enhance ligation efficiency for our lower yielding substrates.

Ligation of synthetic molecules has proven very useful in the study of protein function, localization, and folding. We plan on using the AaT ligation technique to access protein folding studies that have yet to be achieved without significant peptide synthesis. For example, through the demonstrations of model proteins α-synuclein and calmodulin, we have shown that it is possible to incorporate a small synthetic moiety while maximizing the use of expressed protein fragments. AaT N-terminal modifications allow for the accessibility of click reactions at sidechain R-groups and NCL reactions at new residue locations, arginine and lysine, instead of cysteine. Additionally, AaT-ligated homocysteine can be masked, under mild conditions, again allowing for an even greater number of possible ligation sites within proteins. With the increased number of

allowable ligation sites for AaT, small synthetic moieties to study protein structure and function can be inserted into the structure using AaT, with minimal perturbation since the ligation site results in a natural amide bond. The AaT ligation will expand the accessibility of protein structure and folding analysis by incorporating small FRET–based probes such as a thioamide and *p*-cyanophenylalanine pair. The AaT ligation has expanded the opportunities for scientists as a new tool for studying proteins and designing new proteins through protein engineering.

BIBLIOGRAPHY

1. Zimmer, C., On the origin of life on earth. *Science* **2009,** *323* (5911), 198-199.

2. Keasling, J. D.; Mendoza, A.; Baran, P. S., Synthesis: A constructive debate. *Nature* **2012,** *492* (7428), 188-189.

3. Blundell, T. L.; Elliott, G.; Gardner, S. P.; Hubbard, T.; Islam, S.; Johnson, M.; Mantafounis, D.; Murray-Rust, P.; Overington, J.; Pitts, J. E.; Sali, A.; Sibanda, B. L.; Singh, J.; Sternberg, M. J. E.; Sutcliffe, M. J.; Thornton, J. M.; Travers, P., Protein engineering and design. *Philos. Trans. R. Soc. Lond., B.* **1989,** *324* (1224), 447-460.

4. Reetz, M. T., The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed.* **2013,** *52* (10), 2658-2666.

5. Burcu Turanli-Yildiz, C. A., and Z. Petek Cakar, *Chapter 2: Protein engineering methods and applications*. Protein Engineering. 2012 ed.; 2012; p 33-58.

6. Dill, K. A.; MacCallum, J. L., The protein-folding problem, 50 years on. *Science* **2012,** *338* (6110), 1042-1046.

7. Foo, J. L.; Ching, C. B.; Chang, M. W.; Leong, S. S. J., The imminent role of protein engineering in synthetic biology. *Biotechnol. Adv.* **2012,** *30* (3), 541-549.

8. Pellois, J.-P.; Muir, T. W., Semisynthetic proteins in mechanistic studies: Using chemistry to go where nature can't. *Curr. Opin. Chem. Biol.* **2006,** *10* (5), 487-491.

9. Kent, S. B. H., Total chemical synthesis of proteins. *Chem. Soc. Rev.* **2009,** *38* (2), 338-351.

10. Borgia, J. A.; Fields, G. B., Chemical synthesis of proteins. *Trends Biotechnol.* **2000,** *18* (6), 243-251.

11. Boerema, D. J.; Tereshko, V. A.; Kent, S. B. H., Total synthesis by modern chemical ligation methods and high resolution (1.1 Å) x-ray structure of ribonuclease a. *Peptide Science* **2008,** *90* (3), 278-286.

12. Goldberg, J. M.; Batjargal, S.; Petersson, E. J., Thioamides as fluorescence quenching probes: Minimalist chromophores to monitor protein dynamics. *J. Am. Chem. Soc.* **2010,** *132* (42), 14718-14720.

13. Bruice, P. Y., *Organic chemistry*. 3rd ed.; Prentice Hall: 2000.

14. Lehninger, A.; Nelson, D.; Cox, M., *Lehninger Principles of biochemistry*. W. H. Freeman: 2008.

15. Schnolzer, M.; Kent, S. B. H., Constructing proteins by dovetailing unprotected synthetic peptides - backbone-engineered hiv protease. *Science* **1992,** *256* (5054), 221-225.

16. Hackenberger, C. P. R.; Schwarzer, D., Chemoselective ligation and modification strategies for peptides and proteins. *Angew. Chem. Int. Ed.* **2008,** *47* (52), 10030-10074.

17. Johnson, E. C. B.; Kent, S. B. H., Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* **2006,** *128* (20), 6640-6646.

18. Dawson, P. E.; Kent, S. B. H., Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem* **2000,** *69*, 923-960.

19. Flavell, R. R.; Muir, T. W., Expressed protein ligation (epl) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **2009,** *42* (1), 107-116.

20. Liu, X.-Q., Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu. Rev. Genet.* **2000,** *34* (1), 61-76.

21. Perler, F. B.; Davis, E. O.; Dean, G. E.; Gimble, F. S.; Jack, W. E.; Neff, N.; Noren, C. J.; Thorner, J.; Belfort, M., Protein splicing elements - inteins and

exteins - a definition of terms and recommended nomenclature. *Nucleic Acids Res.* **1994,** *22* (7), 1125-1127.

22. Chong, S.; Shao, Y.; Paulus, H.; Benner, J.; Perler, F. B.; Xu, M.-Q., Protein splicing involving the saccharomyces cerevisiae vma intein: The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J. Biol. Chem.* **1996,** *271* (36), 22159-22168.

23. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem* **2003,** *72*, 249-289.

24. Tam, A.; Raines, R. T., Chapter 2 protein engineering with the traceless Staudinger ligation. *Methods enzymol.*, Academic Press: 2009; Vol. Volume 462, pp 25-44.

25. Nilsson, B. L.; Kiessling, L. L.; Raines, R. T., Staudinger ligation: A peptide from a thioester and azide. *Org. Lett.* **2000,** *2* (13), 1939-1941.

26. Tam, A.; Soellner, M. B.; Raines, R. T., Water-soluble phosphinothiols for traceless Staudinger ligation and integration with expressed protein ligation. *J. Am. Chem. Soc.* **2007,** *129* (37), 11421-11430.

27. Prescher, J. A.; Bertozzi, C. R., Chemistry in living systems. *Nat. Chem. Biol.* **2005,** *1* (1), 13-21.

28. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified staudinger reaction. *Science* **2000,** *287* (5460), 2007-2010.

29. Pusterla, I.; Bode, J. W., The mechanism of the α-ketoacid-hydroxylamine amide-forming ligation. *Angew. Chem. Int. Ed.* **2011,** *51* (2), 513-516.

30. Bode, J. W.; Fox, R. M.; Baucom, K. D., Chemoselective amide ligations by decarboxylative condensations of n-alkylhydroxylamines and alpha-ketoacids. *Angew. Chem. Int. Ed.* **2006,** *45* (8), 1248-1252.

31. Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A., Engineering subtilisin and its substrates for efficient ligation of peptide-bonds in aqueous-solution. *Biochemistry* **1991,** *30* (17), 4151-4159.

32. Wehofsky, N.; Koglin, N.; Thust, S.; Bordusa, F., Reverse proteolysis promoted by in situ generated peptide ester fragments. *J. Am. Chem. Soc.* **2003,** *125* (20), 6126-6133.

33. Drenth, J.; Hol, W. G. J.; Jansonius, J. N.; Koekoek, R., A comparison of the three-dimensional structures of subtilisin bpn′ and subtilisin novo. *Cold Spring Harb. Symp. Quant. Biol.* **1972,** *36*, 107-116.

34. Jewett, J. C.; Bertozzi, C. R., Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **2010,** *39* (4), 1272-1279.

35. Tornøe, C. W.; Christensen, C.; Meldal, M., Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem* **2002,** *67* (9), 3057-3064.

36. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise huisgen cycloaddition process: Copper(i)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **2002,** *41* (14), 2596-2599.

37. Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R., A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **2006,** *1* (10), 644-648.

38. Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B., N-terminal protein modification through a biomimetic transamination reaction. *Angew. Chem. Int. Ed.* **2006,** *45* (32), 5307-5311.

39. Mogk, A.; Schmidt, R.; Bukau, B., The n-end rule pathway for regulated proteolysis: Prokaryotic and eukaryotic strategies. *Trends Cell Biol.* **2007,** *17* (4), 165-172.

40. Varshavsky, A., The n-end rule at atomic resolution. *Nat. Struct. Mol. Biol.* **2008,** *15* (12), 1238-1240.

41. Gottesman, S.; Maurizi, M. R., Regulation by proteolysis - energy-dependent proteases and their targets. *Microbiol. Rev.* **1992,** *56* (4), 592-621.

42. Tobias, J. W.; Shrader, T. E.; Rocap, G.; Varshavsky, A., The n-end rule in bacteria. *Science* **1991,** *254* (5036), 1374-1377.

43. Watanabe, K.; Toh, Y.; Suto, K.; Shimizu, Y.; Oka, N.; Wada, T.; Tomita, K., Protein-based peptide-bond formation by aminoacyl-trna protein transferase. *Nature* **2007,** *449* (7164), 867-871.

44. Erbse, A.; Schmidt, R.; Bornemann, T.; Schneider-Mergener, J.; Mogk, A.; Zahn, R.; Dougan, D. A.; Bukau, B., Clps is an essential component of the n-end rule pathway in escherichia coli. *Nature* **2006,** *439* (7077), 753-756.

45. Hanson, P. I.; Whiteheart, S. W., Aaa+ proteins: Have engine, will work. *Nature Rev. Mol. Cell Biol.* **2005,** *6* (7), 519-529.

46. Snider, J.; Thibault, G.; Houry, W. A., The aaa plus superfamily of functionally diverse proteins. *Genome Biol.* **2008,** *9* (4).

47. Varshavsky, A., The n-end rule pathway and regulation by proteolysis. *Protein Sci.* **2011,** *20* (8), 1298-1345.

48. Wang, J. M.; Hartling, J. A.; Flanagan, J. M., The structure of clpp at 2.3 angstrom resolution suggests a model for atp-dependent proteolysis. *Cell* **1997,** *91* (4), 447-456.

49. Thompson, M. W.; Maurizi, M. R., Activity and specificity of escherichia coli clpap protease in cleaving model peptide substrates. *J. Biol. Chem.* **1994,** *269* (27), 18201-18208.

50. Graciet, E.; Hu, R. G.; Piatkov, K.; Rhee, J. H.; Schwarz, E. M.; Varshavsky, A., Aminoacyl-transferases and the n-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (9), 3078-3083.

51. Tasaki, T.; Sriram, S. M.; Park, K. S.; Kwon, Y. T., The n-end rule pathway. In *Annu. Rev. Biochem., vol 81*, Kornberg, R. D., Ed. 2012; Vol. 81, pp 261-289.

52. Varshavsky, A., The n-end rule: Functions, mysteries, uses. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93* (22), 12142-12149.

53. Kaji, A.; Kaji, H.; Novelli, G. D., Soluble amino acid-incorporating system .I. Preparation of system and nature of reaction. *J. Biol. Chem.* **1965,** *240* (3), 1185-1191.

54. Leibowit.Mj; Soffer, R. L., Enzymatic modification of proteins .3. Purification and properties of a leucyl, phenylalanyl transfer ribonucleic acid-protein transferase from escherichia-coli. *J. Biol. Chem.* **1970,** *245* (8), 2066-&.

55. Soffer, R. L., Peptide acceptors in leucine, phenylalanine transfer-reaction. *J. Biol. Chem.* **1973,** *248* (24), 8424-8428.

56. Suto, K.; Shimizu, Y.; Watanabe, K.; Ueda, T.; Fukai, S.; Nureki, O.; Tomita, K., Crystal structures of leucyl/phenylalanyl-trna-protein transferase and its complex with an aminoacyl-trna analog. *EMBO J.* **2006,** *25* (24), 5942-5950.

57. Biarrotte-Sorin, S.; Maillard, A. P.; Delettre, J.; Sougakoff, W.; Arthur, M.; Mayer, C., Crystal structures of weisselia viridescens femx and its complex with udp-murnac-pentapeptide: Insights into femabx family substrates recognition. *Structure* **2004,** *12* (2), 257-267.

58. Scarpulla, R. C.; Deutch, C. E.; Soffer, R. L., Transfer of methionyl residues by leucyl, phenylalanyl-transfer-rna-protein transferase. *Biochem. Biophys. Res. Commun.* **1976,** *71* (2), 584-589.

59. Fung, A. W.; Ebhardt, H. A.; Abeysundara, H.; Moore, J.; Xu, Z.; Fahlman, R. P., An alternative mechanism for the catalysis of peptide bond formation by l/f transferase: Substrate binding and orientation. *J. Mol. Biol.* **2011,** *409* (4), 617-629.

60. Gromadski, K. B.; Rodnina, M. V., Kinetic determinants of high-fidelity trna discrimination on the ribosome. *Mol. Cell* **2004,** *13* (2), 191-200.

61. Effraim, P. R.; Wang, J.; Englander, M. T.; Avins, J.; Leyh, T. S.; Gonzalez, R. L.; Cornish, V. W., Natural amino acids do not require their native trnas for efficient selection by the ribosome. *Nat. Chem. Biol.* **2009,** *5* (12), 947-953.

62. Taki, M.; Kuno, A.; Matoba, S.; Kobayashi, Y.; Futami, J.; Murakami, H.; Suga, H.; Taira, K.; Hasegawa, T.; Sisido, M., Leucyl/phenylalanyl-trna-protein transferase-mediated chemoenzymatic coupling of n-terminal arg/lys units in posttranslationally processed proteins with non-natural amino acids. *Chembiochem* **2006,** *7* (11), 1676-1679.

63. Taki, M.; Kuroiwa, H.; Sisido, M., Chemoenzymatic transfer of fluorescent non-natural amino acids to the n terminus of a protein/peptide. *Chembiochem* **2008,** *9* (5), 719-722.

64. Abramochkin, G.; Shrader, T. E., Aminoacyl-trna recognition by the leucyl/phenylalanyl-trna-protein transferase. *J. Biol. Chem.* **1996,** *271* (37), 22901-22907.

65. Connor, R. E.; Piatkov, K.; Varshavsky, A.; Tirrell, D. A., Enzymatic n-terminal addition of noncanonical amino acids to peptides and proteins. *Chembiochem* **2008,** *9* (3), 366-369.

66. Wagner, A. M.; Fegley, M. W.; Warner, J. B.; Grindley, C. L. J.; Marotta, N. P.; Petersson, E. J., N-terminal protein modification using simple aminoacyl transferase substrates. *J. Am. Chem. Soc.* **2011,** *133* (38), 15139-15147.

67. Wu, P. G.; Brand, L., N-terminal modification of proteins for fluorescence measurements. In *Fluorescence spectroscopy*, 1997; Vol. 278, pp 321-330.

68. Harris, J. M.; Chess, R. B., Effect of pegylation on pharmaceuticals. *Nature Reviews Drug Discovery* **2003,** *2* (3), 214-221.

69. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew. Chem. Int. Ed.* **2005,** *44* (1), 34-66.

70. Basle, E.; Joubert, N.; Pucheault, M., Protein chemical modification on endogenous amino acids. *Chem. Biol.* **2010,** *17* (3), 213-227.

71. Canalle, L. A.; Lowik, D.; van Hest, J. C. M., Polypeptide-polymer bioconjugates. *Chem. Soc. Rev.* **2010,** *39* (1), 329-353.

72. Alouani, S.; Gaertner, H. F.; Mermod, J. J.; Power, C. A.; Bacon, K. B.; Wells, T. N. C.; Proudfoot, A. E. I., A fluorescent interleukin-8 receptor probe produced by targeted labeling at the amino-terminus. *Eur. J. Biochem.* **1995,** *227* (1-2), 328-334.

73. Kinstler, O. B.; Brems, D. N.; Lauren, S. L.; Paige, A. G.; Hamburger, J. B.; Treuheit, M. J., Characterization and stability of n-terminally pegylated rhg-csf. *Pharm. Res.* **1996,** *13* (7), 996-1002.

74. Gale, T. F.; Gorlitzer, J.; O'Brien, S. W.; Williams, D. H., The synthesis and binding of n-terminal derivatives of vancomycin to a bacterial cell wall analogue. *J. Am. Chem. Soc.-Perkin Transactions 1* **1999,** (16), 2267-2270.

75. Ramachandiran, V.; Willms, C.; Kramer, G.; Hardesty, B., Fluorophores at the n terminus of nascent chloramphenicol acetyltransferase peptides affect translation and movement through the ribosome. *J. Biol. Chem.* **2000,** *275* (3), 1781-1786.

76. Jones, D. S.; Cockerill, K. A.; Gamino, C. A.; Hammaker, J. R.; Hayag, M. S.; Iverson, G. M.; Linnik, M. D.; McNeeley, P. A.; Tedder, M. E.; Ton-Nu, H. T.; Victoria, E. J., Synthesis of ljp 993, a multivalent conjugate of the n-terminal domain of beta(2)gpi and suppression of an anti-beta(9)2gpi immune response. *Bioconjugate Chem.* **2001,** *12* (6), 1012-1020.

77. Kinstler, O.; Molineux, G.; Treuheit, M.; Ladd, D.; Gegg, C., Mono-n-terminal poly(ethylene glycol)-protein conjugates. *Advanced Drug Delivery Reviews* **2002,** *54* (4), 477-485.

78. Chelius, D.; Shaler, T. A., Capture of peptides with n-terminal serine and threonine: A sequence-specific chemical method for peptide mixture simplification. *Bioconjugate Chem.* **2003,** *14* (1), 205-211.

79. Arduini, R. M.; Li, Z. F.; Rapoza, A.; Gronke, R.; Hess, D. M.; Wen, D. Y.; Miatkowski, K.; Coots, C.; Kaffashan, A.; Viseux, N.; Delaney, J.; Domon, B.; Young, C. N.; Boynton, R.; Chen, L. L.; Chen, L. Q.; Betzenhauser, M.; Miller, S.; Gill, A.; Pepinsky, R. B.; Hochman, P. S.; Baker, D. P., Expression, purification, and characterization of rat interferon-beta, and preparation of an n-

terminally pegylated form with improved pharmacokinetic parameters. *Protein Expr. Purif.* **2004,** *34* (2), 229-242.

80. Mamaev, S.; Olejnik, J.; Olejnik, E. K.; Rothschild, K. J., Cell-free n-terminal protein labeling using initiator suppressor trna. *Anal. Biochem.* **2004,** *326* (1), 25-32.

81. Baker, D. P.; Lin, E. Y.; Lin, K.; Pellegrini, M.; Petter, R. C.; Chen, L. L.; Arduini, R. M.; Brickelmaier, M.; Wen, D. Y.; Hess, D. M.; Chen, L. Q.; Grant, D.; Whitty, A.; Gill, A.; Lindner, D. J.; Pepinsky, R. B., N-terminally pegylated human interferon-beta-1a with improved pharmacokinetic properties and in vivo efficacy in a melanoma angiogenesis model. *Bioconjugate Chem.* **2006,** *17* (1), 179-188.

82. Scheck, R. A.; Francis, M. B., Regioselective labeling of antibodies through n-terminal transamination. *ACS Chem. Biol.* **2007,** *2* (4), 247-251.

83. Merkel, L.; Beckmann, H. S. G.; Wittmann, V.; Budisa, N., Efficient n-terminal glycoconjugation of proteins by the n-end rule. *Chembiochem* **2008,** *9* (8), 1220-1224.

84. Sharon, J. L.; Puleo, D. A., The use of n-terminal immobilization of pth(1-34) on plga to enhance bioactivity. *Biomaterials* **2008,** *29* (21), 3137-3142.

85. Ebhardt, H. A.; Xu, Z. Z.; Fung, A. W.; Fahlman, R. P., Quantification of the post-translational addition of amino acids to proteins by maldi-tof mass spectrometry. *Anal. Chem.* **2009,** *81* (5), 1937-1943.

86. Gao, W. P.; Liu, W. G.; Mackay, J. A.; Zalutsky, M. R.; Toone, E. J.; Chilkoti, A., In situ growth of a stoichiometric peg-like conjugate at a protein's n-terminus with significantly improved pharmacokinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (36), 15231-15236.

87. Sayers, C. T.; Mantovani, G.; Ryan, S. M.; Randev, R. K.; Keiper, O.; Leszczyszyn, O. I.; Blindauer, C.; Brayden, D. J.; Haddleton, D. M., Site-specific n-terminus conjugation of poly(mpeg(1100)) methacrylates to salmon calcitonin: Synthesis and preliminary biological evaluation. *Soft Matter* **2009,** *5* (16), 3038-3046.

88. Xiao, J. P.; Tolbert, T. J., Synthesis of n-terminally linked protein dimers and trimers by a combined native chemical ligation-cuaac click chemistry strategy. *Org. Lett.* **2009,** *11* (18), 4144-4147.

89. Jia, J.; Chen, W.; Ma, H. M.; Wang, K.; Zhao, C. A., Use of a rhodamine-based bifunctional probe in n-terminal specific labeling of thermomyces lanuginosus xylanase. *Mol. Biosyst.* **2010,** *6* (10), 1829-1833.

90. Wang, J.; Hu, T.; Liu, Y. D.; Zhang, G. F.; Ma, G. H.; Su, Z. G., Kinetic and stoichiometric analysis of the modification process for n-terminal pegylation of staphylokinase. *Anal. Biochem.* **2010,** *412* (1), 114-116.

91. Wildes, D.; Wells, J. A., Sampling the n-terminal proteome of human blood. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (10), 4561-4566.

92. Wu, J. J.; Peng, H. T.; Shek, P. N., Terminal-specific pegylation of polypeptides in a dilute solution. *J. Appl. Polym. Sci.* **2010,** *118* (6), 3269-3273.

93. Dixon, H. B. F.; Weitkamp, L. R., Conversion of n-terminal serine residue of corticotrophin into glycine. *Biochem. J* **1962,** *84* (3), 462-468.

94. Dixon, H. B. F., Transamination of peptides. *Biochem. J* **1964,** *92* (3), 661-666.

95. Geoghegan, K. F.; Ybarra, D. M.; Feeney, R. E., Reversible reductive alkylation of amino-groups in proteins. *Biochemistry* **1979,** *18* (24), 5392-5399.

96. Dixon, H. B. F., N-terminal modification of proteins - a review. *J. Protein Chem.* **1984,** *3* (1), 99-108.

97. Acharya, A. S.; Manjula, B. N., Dihydroxypropylation of amino-groups of proteins - use of glyceraldehyde as a reversible agent for reductive alkylation. *Biochemistry* **1987,** *26* (12), 3524-3530.

98. Qasmi, D.; Rene, L.; Badet, B., Synthesis of cho-co-peptides by n-terminal acylation with a glyoxylyl equivalent. *Tetrahedron Lett.* **1994,** *35* (25), 4343-4344.

99. Li, X. F.; Zhang, L. S.; Hall, S. E.; Tam, J. P., A new ligation method for n-terminal tryptophan-containing peptides using the pictet-spengler reaction. *Tetrahedron Lett.* **2000,** *41* (21), 4069-4073.

100. Scheck, R. A.; Dedeo, M. T.; Lavarone, A. T.; Francis, M. B., Optimization of a biomimetic transamination reaction. *J. Am. Chem. Soc.* **2008,** *130* (35), 11762-11770.

101. Witus, L. S.; Moore, T.; Thuronyi, B. W.; Esser-Kahn, A. P.; Scheck, R. A.; Iayarone, A. T.; Francis, M. B., Identification of highly reactive sequences for plp-mediated bioconjugation using a combinatorial peptide library. *J. Am. Chem. Soc.* **2010,** *132* (47), 16812-16817.

102. Kuhl, P.; Konnecke, A.; Doring, G.; Daumer, H.; Jakubke, H. D., Enzyme-catalyzed peptide-synthesis in biphasic aqueous-organic systems. *Tetrahedron Lett.* **1980,** *21* (10), 893-896.

103. Jakubke, H. D.; Kuhl, P.; Konnecke, A., Basic principles of protease-catalyzed peptide-bond formation. *Angew. Chem. Int. Ed.* **1985,** *24* (2), 85-93.

104. Schellenberger, V.; Jakubke, H. D., Protease-catalyzed kinetically controlled peptide-synthesis. *Angew. Chem. Int. Ed.* **1991,** *30* (11), 1437-1449.

105. Chang, T. K.; Jackson, D. Y.; Burnier, J. P.; Wells, J. A., Subtiligase - a tool for semisynthesis of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91* (26), 12544-12548.

106. Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A., A designed peptide ligase for total synthesis of ribonuclease-a with unnatural catalytic residues. *Science* **1994,** *266* (5183), 243-247.

107. Bordusa, F.; Ullmann, D.; Elsner, C.; Jakubke, H. D., Substrate mimetic mediated peptide synthesis: An irreversible ligation strategy that is independent of substrate specificity. *Angew. Chem. Int. Ed.* **1997,** *36* (22), 2473-2475.

108. Braisted, A. C.; Judice, J. K.; Wells, J. A., Synthesis of proteins by subtiligase. In *Solid-phase peptide synthesis*, 1997; Vol. 289, pp 298-313.

109. Atwell, S.; Wells, J. A., Selection for improved subtiligases by phage display. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (17), 9497-9502.

110. Bordusa, F., Proteases in organic synthesis. *Chem. Rev.* **2002,** *102* (12), 4817-4867.

111. Tolbert, T. J.; Wong, C. H., New methods for proteomic research: Preparation of proteins with n-terminal cysteines for labeling and conjugation. *Angew. Chem. Int. Ed.* **2002,** *41* (12), 2171-2174.

112. Gentle, I. E.; De Souza, D. P.; Baca, M., Direct production of proteins with n-terminal cysteine for site-specific conjugation. *Bioconjugate Chem.* **2004,** *15* (3), 658-663.

113. Yoshihara, H. A. I.; Mahrus, S.; Wells, J. A., Tags for labeling protein n-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* **2008,** *18* (22), 6000-6003.

114. Mazmanian, S. K.; Liu, G.; Hung, T. T.; Schneewind, O., Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **1999,** *285* (5428), 760-763.

115. Mao, H. Y.; Hart, S. A.; Schink, A.; Pollok, B. A., Sortase-mediated protein ligation: A new method for protein engineering. *J. Am. Chem. Soc.* **2004,** *126* (9), 2670-2671.

116. Tanaka, T.; Kamiya, N.; Nagamune, T., N-terminal glycine-specific protein conjugation catalyzed by microbial transglutaminase. *FEBS Lett.* **2005,** *579* (10), 2092-2096.

117. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L., Sortagging: A versatile method for protein labeling. *Nat. Chem. Biol.* **2007,** *3* (11), 707-708.

118. Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M., Site-specific modification and pegylation of pharmaceutical proteins mediated by transglutaminase. *Adv. Drug Deliv. Rev.* **2008,** *60* (1), 13-28.

119. Heal, W. P.; Wickramasinghe, S. R.; Leatherbarrow, R. J.; Tate, E. W., N-myristoyl transferase-mediated protein labelling in vivo. *Org. Biomol. Chem.* **2008,** *6* (13), 2308-2315.

120. Tsukiji, S.; Nagamune, T., Sortase-mediated ligation: A gift from gram-positive bacteria to protein engineering. *Chembiochem* **2009,** *10* (5), 787-798.

121. Nelson, J. W.; Chamessian, A. G.; McEnaney, P. J.; Murelli, R. P.; Kazmierczak, B. I.; Spiegel, D. A., Correction to a biosynthetic strategy for re-engineering the staphylococcus aureus cell wall with non-native small molecules. *ACS Chem. Biol.* **2011,** *6* (9), 971-971.

122. Lahoud, G.; Hou, Y. M., Biosynthesis a new (old) way of hijacking trna. *Nat. Chem. Biol.* **2010,** *6* (11), 795-796.

123. Kaji, A.; Novelli, G. D.; Kaji, H., A soluble amino acid incorporating system. *Biochem. Biophys. Res. Commun.* **1963,** *10* (5), 406-409.

124. Schuenemann, V. J.; Kralik, S. M.; Albrecht, R.; Spall, S. K.; Truscott, K. N.; Dougan, D. A.; Zeth, K., Structural basis of n-end rule substrate recognition in escherichia coli by the clpap adaptor protein clps. *Embo Reports* **2009,** *10* (5), 508-514.

125. Kaji, A.; Kaji, H.; Novelli, G. D., Soluble amino acid-incorporating system .2. Soluble nature of system and characterization of radioactive product. *J. Biol. Chem.* **1965,** *240* (3), 1192-1197.

126. M., L.; Soffer, R. L., A soluble enzyme from escherichia coli which catalyzes transfer of leucine and phenylalanine from trna to acceptor proteins. *Biochem. Biophys. Res. Commun.* **1969,** *36* (1), 47-53.

127. Leibowitz, M. J., Enzymatic modification of proteins .7. Substrate specificity of leucyl,phenylalanyl-transfer ribonucleic acid-protein transferase. *J. Biol. Chem.* **1971,** *246* (17), 5207-5212.

128. Taki, M.; Sisido, M., Leucyl/phenylalanyl(l/f)-trna-protein transferase-mediated aminoacyl transfer of a nonnatural amino acid to the n-terminus of peptides and

proteins and subsequent functionalization by bioorthogonal reactions. *Biopolymers* **2007,** *88* (2), 263-271.

129. Ebisu, K.; Tateno, H.; Kuroiwa, H.; Kawakami, K.; Ikeuchi, M.; Hirabayashi, J.; Sisido, M.; Taki, M., N-terminal specific point-immobilization of active proteins by the one-pot next-a method. *Chembiochem* **2009,** *10* (15), 2460-2464.

130. Abramochkin, G.; Shrader, T. E., The leucy/phenylalanyl-transfer-rna-protein transferase - overexpression and characterization of substrate recognition, domain-structure, and secondary structure. *J. Biol. Chem.* **1995,** *270* (35), 20621-20628.

131. Fersht, A., *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. 3rd ed.; W. H. Freeman: 1998.

132. Mercier, J. C.; Grosclau, F.; Ribadeau, B., Primary structure of bovine alphas1 casein - complete sequence. *Eur. J. Biochem.* **1971,** *23* (1), 41-51.

133. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **2001,** *40* (11), 2004-2021.

134. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G., Bioconjugation by copper(i)-catalyzed azide-alkyne [3+2] cycloaddition. *J. Am. Chem. Soc.* **2003,** *125* (11), 3192-3193.

135. Huisgen, R., Kinetics and reaction mechanisms: Selected examples from the experience of forty years. *Pure App Chem.* **1989,** *61*, 613–628.

136. Meldal M, T. C., Cu-catalyzed azide-alkyne cycloaddition. *Chem Rev.* **2008,** *108*, 2952–3015.

137. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **2009,** *48* (38), 6974-6998.

138. Carrico, I. S., Chemoselective modification of proteins: Hitting the target. *Chem. Soc. Rev.* **2008,** *37* (7), 1423-1431.

139. Ogilvie, K. K.; l., L. R., Use of isobutyloxycarbonyl as a blocking group in preparation of 3'-o-p-monomethoxytritylthymidine. *J. Org. Chem.* **1967,** *32* (7), 2365-2366.

140. Taiji, M.; Yokoyama, S.; Miyazawa, T., Transacylation rates of (aminoacyl)adenosine moiety at the 3'-terminus of aminoacyl transfer ribonucleic-acid. *Biochemistry* **1983,** *22* (13), 3220-3225.

141. Bradford, M. M., Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* **1976,** *72* (1-2), 248-254.

142. Ellman J, M. D., Anthonycahill S, Noren CJ, Schultz PG., Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods Enzymol.* **1991,** *202*, 301–336.

143. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G., Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed.* **2009,** *48* (52), 9879-9883.

144. Muralidharan, V.; Muir, T. W., Protein ligation: An enabling technology for the biophysical analysis of proteins. *Nature Methods* **2006,** *3* (6), 429-438.

145. Frederick Neidhardt, J. I., and Moselio Schaechte, *Physiology of the bacterial cell: A molecular approach* Sinauer Associates, Inc.: Sunderland, MA, 1990; p 507.

146. Okazaki, K.; Yamada, H.; Imoto, T., A convenient s-2-aminoethylation of cysteinyl residues in reduced proteins. *Anal. Biochem.* **1985,** *149* (2), 516-520.

147. Yan, L. Z.; Dawson, P. E., Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **2001,** *123* (4), 526-533.

148. Hopkins, C. E.; Hernandez, G.; Lee, J. P.; Tolan, D. R., Aminoethylation in model peptides reveals conditions for maximizing thiol specificity. *Arch. Biochem. Biophys.* **2005,** *443* (1-2), 1-10.

149. Crich, D.; Banerjee, A., Native chemical ligation at phenylalanine. *J. Am. Chem. Soc.* **2007,** *129* (33), 10064-10065.

150. Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J. L.; Danishefsky, S. J., Native chemical ligation at valine: A contribution to peptide and glycopeptide synthesis. *Angew. Chem. Int. Ed.* **2008,** *47* (44), 8521-8524.

151. Haase, C.; Rohde, H.; Seitz, O., Native chemical ligation at valine. *Angew. Chem. Int. Ed.* **2008,** *47* (36), 6807-6810.

152. Yang, R. L.; Pasunooti, K. K.; Li, F. P.; Liu, X. W.; Liu, C. F., Dual native chemical ligation at lysine. *J. Am. Chem. Soc.* **2009,** *131* (38), 13592-13593.

153. Chen, J.; Wang, P.; Zhu, J. L.; Wan, Q.; Danishefsky, S. J., A program for ligation at threonine sites: Application to the controlled total synthesis of glycopeptides. *Tetrahedron* **2010,** *66* (13), 2277-2283.

154. Harpaz, Z.; Siman, P.; Kumar, K. S. A.; Brik, A., Protein synthesis assisted by native chemical ligation at leucine. *Chembiochem* **2010,** *11* (9), 1232-1235.

155. Townsend, S. D.; Tan, Z. P.; Dong, S. W.; Shang, S. Y.; Brailsford, J. A.; Danishefsky, S. J., Advances in proline ligation. *J. Am. Chem. Soc.* **2012,** *134* (8), 3912-3916.

156. Offer, J.; Boddy, C. N. C.; Dawson, P. E., Extending synthetic access to proteins with a removable acyl transfer auxiliary. *J. Am. Chem. Soc.* **2002,** *124* (17), 4642-4646.

157. Chatterjee, C.; McGinty, R. K.; Pellois, J. P.; Muir, T. W., Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chem. Int. Ed.* **2007,** *46* (16), 2814-2818.

158. Saxon, E.; Armstrong, J. I.; Bertozzi, C. R., A "traceless" staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2000,** *2* (14), 2141-2143.

159. Nilsson, B. L.; Kiessling, L. L.; Raines, R. T., High-yielding staudinger ligation of a phosphinothioester and azide to form a peptide. *Org. Lett.* **2001,** *3* (1), 9-12.

160. Tam, J. P.; Yu, Q. T., Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones. *Biopolymers* **1998,** *46* (5), 319-327.

161. Saporito, A.; Marasco, D.; Chambery, A.; Botti, P.; Monti, S. M.; Pedone, C.; Ruvo, M., The chemical synthesis of the gstl protein by ncl on a x-met site. *Biopolymers* **2006,** *83* (5), 508-518.

162. Pachamuthu, K.; Schmidt, R. R., Synthesis of methionine containing peptides related to native chemical ligation. *Synlett* **2003,** (5), 659-662.

163. Aussedat, B.; Fasching, B.; Johnston, E.; Sane, N.; Nagorny, P.; Danishefsky, S. J., Total synthesis of the alpha-subunit of human glycoprotein hormones: Toward fully synthetic homogeneous human follicle-stimulating hormone. *J. Am. Chem. Soc.* **2012,** *134* (7), 3532-3541.

164. Auluck, P. K.; Caraveo, G.; Lindquist, S., Alpha-synuclein: Membrane interactions and toxicity in parkinson's disease. *Annual Review of Cell and Developmental Biology, Vol 26* **2010,** *26*, 211-233.

165. Chin, D.; Means, A. R., Calmodulin: A prototypical calcium sensor. *Trends Cell Biol.* **2000,** *10* (8), 322-328.

166. Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A., Discovery of aminoacyl-trna synthetase activity through cell-surface display of noncanonical amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (27), 10180-10185.

167. Trujillo, M.; Radi, R., Peroxynitrite reaction with the reduced and the oxidized forms of lipoic acid: New insights into the reaction of peroxynitrite with thiols. *Arch. Biochem. Biophys.* **2002,** *397* (1), 91-98.

168. Vamvaca, K.; Volles, M. J.; Lansbury, P. T., The first n-terminal amino acids of alpha-synuclein are essential for alpha-helical structure formation in vitro and membrane binding in yeast. *J. Mol. Biol.* **2009,** *389* (2), 413-424.

169. Bartels, T.; Ahlstrom, L. S.; Leftin, A.; Kamp, F.; Haass, C.; Brown, M. F.; Beyer, K., The n-terminus of the intrinsically disordered protein alpha-synuclein

triggers membrane binding and helix folding. *Biophys. J.* **2010,** *99* (7), 2116-2124.

170. Bartels, T.; Choi, J. G.; Selkoe, D. J., Alpha-synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **2011,** *477* (7362), 107-110.

171. Vamvaca, K.; Lansbury, P. T.; Stefanis, L., N-terminal deletion does not affect alpha-synuclein membrane binding, self-association and toxicity in human neuroblastoma cells, unlike yeast. *J. Neurochem.* **2011,** *119* (2), 389-397.

172. Fauvet, B.; Fares, M. B.; Samuel, F.; Dikiy, I.; Tandon, A.; Eliezer, D.; Lashuel, H. A., Characterization of semisynthetic and naturally nα-acetylated alpha-synuclein in vitro and in intact cells: Implications for aggregation and cellular properties of alpha-synuclein. *J. Biol. Chem.* **2012**.

173. Batjargal, S.; Wang, Y. J.; Goldberg, J. M.; Wissner, R. F.; Petersson, E. J., Native chemical ligation of thioamide-containing peptides: Development and application to the synthesis of labeled alpha-synuclein for misfolding studies. *J. Am. Chem. Soc.* **2012,** *134* (22), 9172-9182.

174. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped blast and psi-blast: A new generation of protein database search programs. *Nucleic Acids Res.* **1997,** *25* (17), 3389-3402.

175. Dong, X. S.; Kato-Murayama, M.; Muramatsu, T.; Mori, H.; Shirouzu, M.; Bessho, Y.; Yokoyama, S., The crystal structure of leucyl/phenylalanyl-trna-protein transferase from escherichia coli. *Protein Sci.* **2007,** *16* (3), 528-534.

176. Kwon, Y. T.; Kashina, A. S.; Davydov, I. V.; Hu, R. G.; An, J. Y.; Seo, J. W.; Du, F.; Varshavsky, A., An essential role of n-terminal arginylation in cardiovascular development. *Science* **2002,** *297* (5578), 96-99.

177. Wang, J. L.; Han, X. M.; Saha, S.; Xu, T.; Rai, R.; Zhang, F. L.; Wolf, Y. I.; Wolfson, A.; Yates, J. R.; Kashina, A., Arginyltransferase is an atp-independent self-regulating enzyme that forms distinct functional complexes in vivo. *Chem. Biol.* **2011,** *18* (1), 121-130.

178. Wong, C. C. L.; Xu, T.; Rai, R.; Bailey, A. O.; Yates, J. R.; Wolf, Y. I.; Zebroski, H.; Kashina, A., Global analysis of posttranslational protein arginylation. *PLoS Biol.* **2007,** *5* (10), 2231-2242.